Analog Circuit Design

VISIT…

Arthur H. M. van Roermund • Herman Casier
Michiel Steyaert
Editors

Analog Circuit Design

Smart Data Converters, Filters on Chip, Multimode Transmitters

Springer

Editors
Dr. Arthur H. M. van Roermund
Department of Electrical Engineering
Eindhoven University of Technology
5600 MB Eindhoven
Netherlands
a.h.m.v.roermund@tue.nl

Dr. Herman Casier
Avondster 6
8520 Kuurne
Belgium
herman_casier@ieee.org

Prof. Michiel Steyaert
Department of Electrical
Engineering (ESAT)
Katholieke Universiteit Leuven
Kasteelpark Arenberg 10
3001 Leuven
Belgium
michiel.steyaert@esat.kuleuven.be

ISBN 978-90-481-3082-5 e-ISBN 978-90-481-3083-2
DOI 10.1007/978-90-481-3083-2
Springer Dordrecht Heidelberg London New York

Library of Congress Control Number: 2009929389

© Springer Science+Business Media B.V. 2010
No part of this work may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, microfilming, recording or otherwise, without written permission from the Publisher, with the exception of any material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use by the purchaser of the work.

Cover design: eStudio Calamar S.L.

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

This book is part of the Analog Circuit Design series and contains contributions of the speakers of the 18th workshop on Advances in Analog Circuit Design (AACD), which was organized by Sven Mattisson of Ericsson. The workshop was held in Lund, Sweden, from March 31 to April 2, 2009.

The book comprises three parts, covering advanced analog and mixed-signal circuit design fields that are considered as very important by the circuit design community:

- Smart Data Converters
- Filters on Chip
- Multimode Transmitters

Each part is set up with six papers from experts in the field.

The aim of the AACD workshop is to bring together a group of expert designers to discuss new developments and future options. Each workshop is then followed by the publication of a book by Springer in their successful series of Analog Circuit Design. This book is number 18 in this series. The books can be seen as a reference for all people involved in analog and mixed-signal design. The full list of the previous books and topics in the series is given next.

We are confident that this book, like its predecessors, provides a valuable contribution to our analog and mixed-signal circuit-design community.

Arthur van Roermund.

The topics covered before in this series:

2008	Pavia (Italy)	High-speed Clock and Data Recovery
		High-performance Amplifiers
		Power Management
2007	Oostende (Belgium)	Sensors, Actuators and Power Drivers for the Automotive and Industrial Environment
		Integrated PAs from Wireline to RF
		Very High Frequency Front Ends

(continued)

(continued)

2006	Maastricht (The Netherlands)	High-speed AD Converters
		Automotive Electronics: EMC Issues
		Ultra Low Power Wireless
2005	Limerick (Ireland)	RF Circuits: Wide Band, Front-Ends, DACs
		Design Methodology and Verification of RF and Mixed-Signal Systems
		Low Power and Low Voltage
2004	Montreux (Swiss)	Sensor and Actuator Interface Electronics Integrated
		High-Voltage Electronics and Power Management
		Low-Power and High-Resolution ADCs
2003	Graz (Austria)	Fractional-N Synthesizers Design for Robustness
		Line and Bus drivers
2002	Spa (Belgium)	Structured Mixed-Mode Design
		Multi-Bit Sigma-Delta Converters
		Short-Range RF Circuits
2001	Noordwijk (The Netherlands)	Scalable Analog Circuits
		High-Speed D/A Converters
		RF Power Amplifiers
		High-Speed A/D Converters
2000	Munich (Germany)	Mixed-Signal Design PLLs and Synthesizers
1999	Nice (France)	XDSL and other Communication Systems
		RF-MOST Models and Behavioural Modelling
		Integrated Filters and Oscillators
1998	Copenhagen (Denmark)	1-Volt Electronics
		Mixed-Mode Systems
		LNAs and RF Power Amps for Telecom
1997	Como (Italy)	RF A/D Converters
		Sensor and Actuator Interfaces
		Low-Noise Oscillators, PLLs and Synthesizers
1996	Lausanne (Swiss)	RF CMOS Circuit Design
		Bandpass Sigma Delta and Other Data Converters
		Translinear Circuits
1995	Villach (Austria)	Low-Noise/Power/Voltage
		Mixed-Mode with CAD Tools
		Voltage, Current and Time References
1994	Eindhoven (Netherlands)	Low-Power Low-Voltage
		Integrated Filters
		Smart Power
1993	Leuven (Belgium)	Mixed-Mode A/D Design
		Sensor Interfaces
		Communication Circuits
1992	Scheveningen (The Netherlands)	OpAmps
		ADC
		Analog CAD

Contents

Part I Smart Data Converters

1 LMS-Based Digital Assisting for Data Converters 3
Bang-Sup Song

2 Pipelined ADC Digital Calibration Techniques and Tradeoffs 23
Imran Ahmed

3 High-Resolution and Wide-Bandwidth CMOS Pipeline AD Converters 43
Hans Van de Vel

4 A Signal Processing View on Time-Interleaved ADCS 61
Christian Vogel

5 DAC Correction and Flexibility, Classification, New Methods and Designs 79
Georgi Radulov, Patrick Quinn, Hans Hegt, and Arthur van Roermund

6 Smart CMOS Current-Steering D/A-Converters for Embedded Applications 107
Martin Clara, Daniel Gruber, and Wolfgang Klatzer

Part II Filters On-Chip

7 Synthesis of Low-Sensitivity Analog Filters 129
Lars Wanhammar

8 High-Performance Continuous-Time Filters with On-Chip Tuning 147
Jose Silva-Martinez and Aydın İ. Karşılayan

9 Source-Follower-Based Continuous Time Analog Filters 167
Stefano D'Amico, Marcello De Matteis, and Andrea Baschirotto

10 Reconfigurable Active-RC Filters with High Linearity and Low Noise for Home Networking Applications 189
Jan Vandenbussche, Jan Crols, and Yuichi Segawa

11 On-Chip Instantaneously Companding Filters for Wireless Communications 203
Vaibhav Maheshwari and Wouter A. Serdijn

12 BAW-IC CO-Integration Tunable Filters at GHz Frequencies 219
Andreia Cathelin, Stéphane Razafimandimby, and Andreas Kaiser

Part III Multi-mode Transmitters

13 Multimode Transmitters: Easier with Strong Nonlinearity.....247
Earl McCune

14 RBS High Efficiency Power Amplifier Research – Challenges and Possibilities 259
Bo Berglund, Ulf Gustavsson, Johan Thorebäck, Thomas Lejon, and Ericsson AB

15 Multi-Mode Transmitters in CMOS.....275
Manel Collados, Xin He, Jan van Sinderen, and Raf Roovers

16 Challenges for Mobile Terminal CMOS Power Amplifiers 295
Patrick Reynaert

17 Multimode Transmitters with ΔΣ-Based All-Digital RF Signal Generation 305
A. Frappé, A. Kaiser, A. Flament, and B. Stefanelli

18 Switched Mode Transmitter Architectures 325
Henrik Sjöland, Carl Bryant, Vandana Bassoo, and Mike Faulkner

Part I
Smart Data Converters

The first part of this book covers the theme 'Smart Data Converters'. As the name indicates, it deals with smart converters that have some kind of smartness implemented on chip, to make the converter better in performance for a given amount of resources like power dissipation and area. On-chip smartness might also result in an increase in yield, a decrease in design effort, a higher flexibility, more functionality and/or broader applicability. All these aspects in turn also pay off in less cost.

The Part starts with AD converters. Three types of AD converters achieve considerable attention nowadays, and are therefore addressed here: pipelined, Sigma-Delta, and time-interleaved AD converters. The first paper discusses both LMS-based calibrated pipeline and Sigma-Delta converters and also makes some comparisons between the two. The second paper fully focuses on pipeline converters and addresses several calibration techniques. The third paper discusses a calibrated pipeline in the application context of a multi-channel, and thus wideband, front end of a cellular base station.

Next we proceed with a paper on time-interleaved converters. Here the problem is in the equality of the channels in terms of gain, time, and more generically seen: in spectral behaviour. This paper will address the problem from a signal-processing point of view, so from a higher level of abstraction, to show what theoretical approaches are possible to correct for lower-level induced channel differences, and what are the tradeoffs between them, on an algorithmic level.

Finally we end up with two DA papers. The first one gives an overview and classification of smart approaches for Current-Steering DAs, as they are known now in literature, shows solutions for missing approaches, and addresses flexibility as one of the features of smart converters. The second DA paper also addresses Current-Steering DAs, but focuses more specifically on the embedding of these kinds of converters in systems-on-chip (SoCs), which implies some extra constraints that should be met.

Arthur van Roermund

Chapter 1
LMS-Based Digital Assisting for Data Converters

Bang-Sup Song

Abstract Aggressive device scaling down to the nano-meter range offers IC designers both opportunities and challenges. Digital designers benefit greatly from the system flexibility and affordability, but analog/RF designers are struggling with flawed devices. Since scaled devices are faster and smaller, the incentive to use such strengths advantageously has prompted many efforts to overcome analog imperfection by digital means. Designers are introducing more DSP functionality to enhance the performance of analog/RF systems. More intelligence is being built into analog/RF designs as in linear PA, RF receiver front-end, ADC/DAC, digital PLL, etc. Such pervasive design techniques with digital assisting will prevail in the future SOC design. After a brief overview of the trend, examples of the LMS-based calibration algorithm applied to the pipeline and CT cascaded $\Delta\Sigma$ modulator are discussed.

1.1 Introduction

CMOS analog design has evolved along with the device scaling for three decades since early 1980s. In its early days, the supply voltage was higher, the opamp had high gain while devices were slow, and the crude lithography limited the capacitor matching only to 8–9 b level. The two-stage opamp and the simple SAR were predominantly used at low 10 s of kHz range mostly for the voice-band processing. The $\Delta\Sigma$ modulator was feasible, but digital filtering was very costly. This changed in 1990s as CMOS was aggressively scaled down towards the sub-micron range. In this middle period, the supply voltage was lowered from high 5–10 to 1.8–3.3 V, and devices were fast enough to digitize the video band and beyond. Two ADC architectures stood out – pipeline for high-speed communications and video, and $\Delta\Sigma$ for high-resolution audio. Cascaded single-stage opamp was adopted, and many ADC calibration techniques were developed to enhance the resolution of the pipelined

B.-S. Song (✉)
Department of Electrical and Computer Engineering, University of California, San Diego, USA
e-mail: song@ece.ucsd.edu

A.H.M. van Roermund et al. (eds.), *Analog Circuit Design: Smart Data Converters, Filters on Chip, Multimode Transmitters*, DOI 10.1007/978-90-481-3083-2_1,
© Springer Science+Business Media B.V. 2010

ADC to above 12 b range. Now in 2000s, CMOS is still being scaled down from the sub-micron to the nano-meter range, and the supply voltage also approaches sub 1 V. The real advantages of such scaled devices are raw speed, fine lithography, and almost free digital circuitry. The fine-line lithography also made the bare capacitor matching of 12 b level feasible.

These days, analog engineers start with faster and more accurate devices than earlier generations did, and most designs turn out to be already high speed and high resolution with low power. However, a couple of problems should be dealt with. With low supply voltages, SNR is limited by the signal swing, and the low gain defeats any design effort to use the conventional analog design wisdom accumulated over decades. In addition, the device leakage makes any accurate switched-capacitor design difficult. In fact, it appears that the analog design trend is reset, and it starts over again from the beginning. Two or multi-stage opamps are back, but their gain is still low and non-linear. Old ADC designs such as algorithmic, SAR, and time-interleaving are also being revisited. In order to avoid using low-gain non-linear opamps, the new breeds of ADC architectures that use no opamps started to emerge. Examples are comparator-based pipeline ADCs and quantizers based on time resolution. On the other hand, the industry has grown with the powerful broadband digital processing that enables SOCs such as cellphone, WiFi, TV tuner, ... This new environment has created a demand for wideband ADCs such as IF quantizers with very high SFDR to facilitate the digital channel filtering after quantizing the desired spectrum with large blocker channels. Also for high-resolution graphic or imaging, high SNR over 80 dB and low-level linearity over 15 b at sampling rates over 50 MS/s are required to resolve even dark images further in more details. It is challenging to meet such demands with scaled low-voltage CMOS.

Two high-resolution ADC architectures that can meet such high demands are the calibrated pipelined ADC and the CT $\Delta\Sigma$ modulator. The former is now well established enough to calibrate even the opamp non-linearity. The latter exhibits many desirable features in wireless applications and gains momentum as it requires no anti-aliasing, and SNR is improved not by the calibration accuracy but by the feedback. In the following sections, after high-resolution ADCs and their fundamental limits are overviewed, an LMS-based resolution-enhancing technique is introduced, which eliminates the residual error after calibration using the zero-forcing LMS servo feedback concept.

1.2 High-Resolution ADCs

High-resolution ADCs sampling at 10–250 MS/s with 12–16 b linearity have been implemented mostly with SAR, $\Delta\Sigma$, or pipeline architectures as shown in the resolution spectrum of Fig. 1.1. The SAR is very desirable for low-voltage and low-power applications since it uses only one comparator. However, the pipeline offers a significant speed advantage while the $\Delta\Sigma$ is more robust in achieving high resolution. High-resolution ADCs at high sampling rates are only feasible with scaled

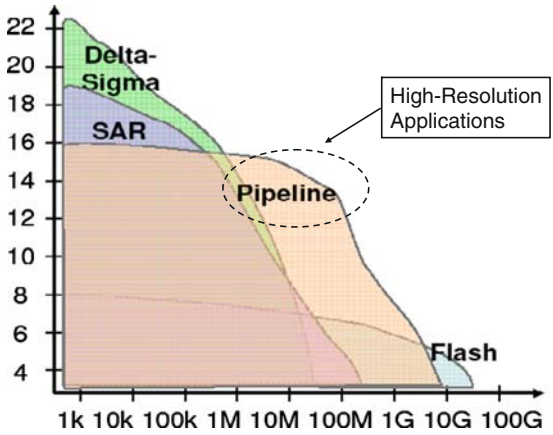


Fig. 1.1 Resolution vs. bandwidth of ADCs

technology with low supply voltages, and their performance is commonly characterized by their linearity measured by SFDR or THD. Such ADCs with high linearity but poor SNR are allowed in systems performing digital filtering.

The earliest effort to enhance the ADC resolution was an EPROM-based code-mapping technique using a radix <2, which warrants monotonicity and proper addressing [1]. However, it was possible only at factory since it required external precision instruments. The first *self-calibration* concept for the SAR was introduced to measure capacitor mismatch errors, to store them digitally, and to subtract them during the normal operation [2, 3]. This self-calibrated SAR based on the charge redistribution capacitor array was slow, and the over-sampling ADC covered the voice or audio band better. Also one critical flaw of the high-resolution SAR was the slowly-varying offset of the comparator due to the stress inflicted upon the input differential pair of the comparator when several decisions are made repeatedly after one input sampling. Finally, the Nyquist-rate ADC above the video band became a reality when the pipelined architecture was introduced [4], and the capacitor-array MDAC as a residue amplifier enabled the development of high-resolution ADCs [5–7]. The switched-capacitor MDAC performs multiple functions of sampling, DAC subtraction, and amplification as a residue amplifier in the pipelined ADC or as an integrator in the DT $\Delta\Sigma$ modulator.

Figure 1.2 compares the switched-capacitor MDAC with the CT integrator. The former is used in an open-ended system, and the residue amplifier should settle with an absolute accuracy. However, the latter rests inside the feedback loop, and its gain and non-linearity errors are reduced by the loop gain. One critical factor to consider at the system level is the anti-aliasing requirement. Nyquist-rate ADCs need high-order anti-aliasing filters when operated at close to the Nyquist rate while CT $\Delta\Sigma$ modulators need no anti-aliasing at all. The speed advantage of the pipelined ADC over the $\Delta\Sigma$ modulator has always been by a factor of 2 to 4, but the gap was quickly narrowed as technology was scaled. A good example is the first digitally-calibrated 1 MS/s, 16 b ADC product (MAX1200) overtaken by the

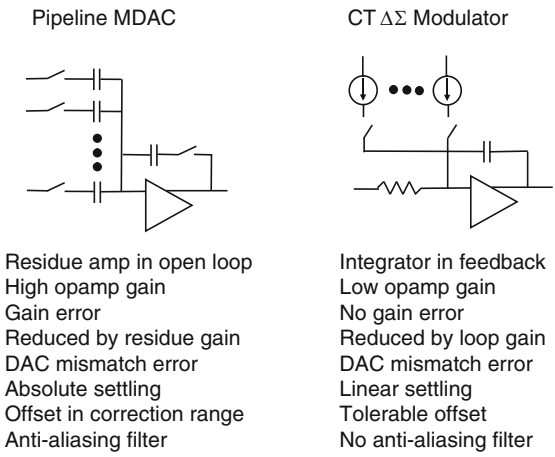


Fig. 1.2 Pipeline vs. CT $\Delta\Sigma$ modulator

$\Delta\Sigma$ ADC [8]. It also happened earlier in 1980s when the $\Delta\Sigma$ modulator replaced the self-calibrated SAR as audio coders. Even today, the same competition between the pipelined ADC and the $\Delta\Sigma$ modulator still continues. The common theme in this competition for the best is now *calibration*. The CT $\Delta\Sigma$ modulator also needs calibration as the over-sampling ratio is lowered to 6–8 approaching the Nyquist rate for high-speed operation.

All earlier calibration was done in the analog domain although measured errors were stored digitally. An effort to perform the error subtraction in the digital domain led to the *digital calibration* concept [9, 10], but error measurements were still performed in a separate measurement cycle. The term such as *foreground* or *background* is used depending on how the error measurement is performed [11–15]. The latest background error measurement technique has evolved into a very sophisticated one, called PN dithering. The PN sequence is a pseudo-random binary pulse sequence with an equal probability of 1 or -1 over a long sample period. It was used for the pulse modulation for the radar jamming during the World War II, and also for the military security communications known as Spread Spectrum and Global Positioning System [16], which are now well known as commercial systems such as CDMA and GPS. The first example of using the PN sequence to enhance the ADC resolution was to dither the ADC for low DNL, and the injected dither was also subtracted digitally [17]. An attempt was made to calibrate the inter-stage gain error by injecting the PN-modulated dither and measuring it digitally [18]. The PN-dithering scheme has been investigated extensively later on [19, 20]. When the term *digital background* is used, errors are measured during the normal operation and calibrated in the digital domain. The digital background calibration is preferred to other high-resolution techniques because it can track long-term process variations, and the digital power and area overhead diminishes as CMOS is scaled down.

1.3 Limits of ADC Resolution

ADC resolution is basically limited by the accuracy of the reference levels used for decision in multiple stages. In the pipelined ADC, it appears as the reference range mismatch between stages, which commonly results from capacitor mismatch and finite opamp gain. In the cascaded CT $\Delta\Sigma$ modulator as will be discussed later, it appears as a time-constant mismatch between analog filter and the digital noise cancellation filter. It is the problem in all multi-step or sub-ranging architectures. In particular, the first-stage residue accuracy limits the ADC performance, and the requirement gets less stringent in later stages because the inter-stage gain of the residue is implemented. However, the accuracy of this inter-stage gain is the fundamental source of the ADC non-linearity measured by DNL and INL.

The left-hand side of Fig. 1.3 shows the 2 b residue amplified by 4 covering $4V_{ref}$ in the pipelined ADC. Since the input range of the later stage is still V_{ref}, the residue output should be fitted into the V_{ref} range. There are two ways of doing it. One is folding, and the other is shifting. If three comparators are placed at $V_{ref}/4$, $2V_{ref}/4$, and $3V_{ref}/4$ as marked by triangles, these out-of-range segments above V_{ref} can be shifted down to fit into the input range of the later stage if V_{ref}, $2V_{ref}$, and $3V_{ref}$ are subtracted as shown. These subtracted analog V_{ref}'s should be restored digitally as shown on the right-hand side by adding digital numbers 01, 10, 11, respectively.

The problem arises when the analog and digital V_{ref}'s are mismatched. The analog V_{ref} is subtracted in the switched-capacitor residue amplifier by flipping the bottom of one unit capacitor to V_{ref} depending on the comparator decision as shown in Fig. 1.4. There can be many error sources in this residue – capacitor mismatch between two equal capacitors, finite opamp gain, and opamp non-linearity. As a result, the analog V_{ref} step does not match the digital V_{ref} step. That is, the digital output can experience a small step discontinuity at major comparator threshold points

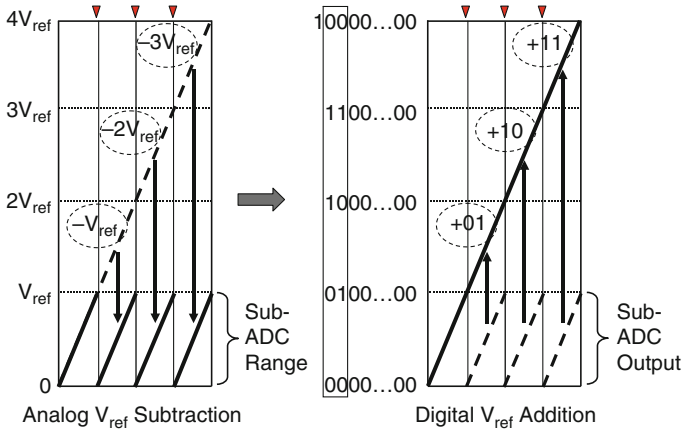


Fig. 1.3 Analog range shift for sub-ranging and its digital restoration

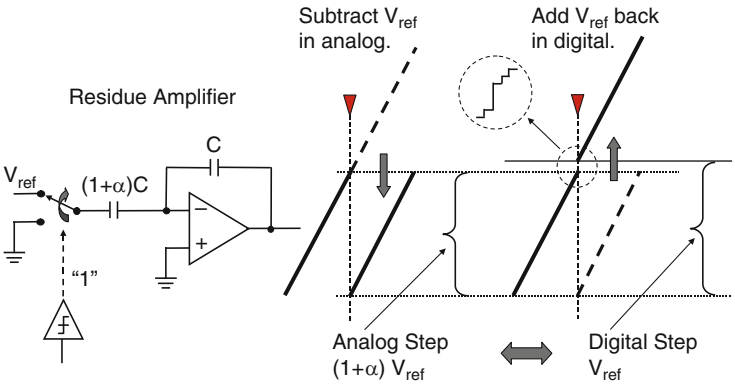


Fig. 1.4 Switched-capacitor V_{ref} subtraction

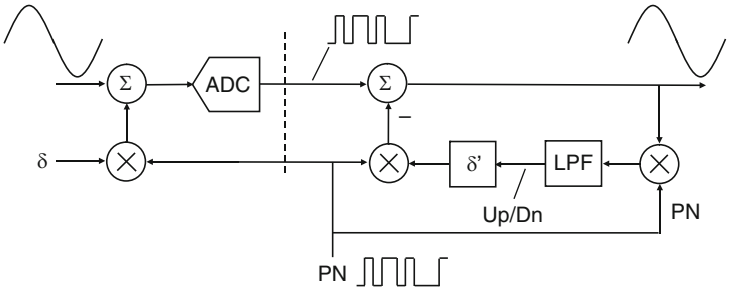


Fig. 1.5 LMS-based calibration with zero-forcing feedback

as circled with the dashed line. If the analog step is smaller than the digital step, missing codes occur. In the standard ADC code-density test, such missing codes are rarely measured since the noise works like dithers and digital codes are spread over the neighboring ones. On the contrary, if larger, the transfer function becomes non-monotonic, which is usually measured by large positive DNL. This reference mismatch is the main source of the DNL and INL errors in all multi-step architectures. The standard digital correction is to restore the subtracted analog V_{ref} with an ideal digital V_{ref}, which is simply a full-range MSB bit. On the other hand, the digital calibration is to restore the subtracted analog V_{ref} with an actual digital step V_{ref}.

1.4 Zero-Forcing LMS Algorithm

Applying the zero-forcing LMS algorithm to enhance the ADC resolution requires the following three steps as shown in Fig. 1.5. First, the gain or DAC error δ should be separated and embedded in the signal after PN-modulated. Second, after the same PN-modulated error δ' is subtracted, the residual error $(\delta - \delta')$ needs to be correlated

using the same PN sequence to determine the sign of the residual error. Third, the residual error is forced to be zero by feedback based on the polarity of the residue error.

This adaptive zero-forcing servo feedback algorithm does not require any specific signal condition. It behaves very similarly to the classic LMS algorithm [21]. The sign–sign algorithm greatly simplifies the digital implementation of the algorithm. The LPF can be implemented with a digital integrate-and-dump *SINC* function with an extremely high over-sampling ratio. Since the error is updated slowly with a negligible step at a time, the stability is not an issue. The sign–sign LMS algorithm has been applied to improve analog performance such as image rejection, spurious fractional tone, and capacitor mismatch [22–25]. When it is applied to the pipelined ADC calibration, the DAC and gain error should be embedded in the large residue output, and correlating only the small PN-modulated error out relies on an assumption that the large residue output averages out to be smaller than the error by 2^{-15} for 15 b, for example. The correlated error term increases linearly as more samples are integrated, but the de-correlated signal term randomly fluctuates. After the error polarity detection, the error subtraction can be done either in the analog or digital domain.

1.5 LMS-Based Calibration of the Pipelined ADC

For any background calibration to be useful, it is necessary that its impact on analog circuits should be minimum, and the calibration cycle should be short. The background calibration by PN-dithering has two constraints. One is the measurement time constraint. With a large un-correlated signal present, it is difficult to detect a small PN-modulated error. In particular, a large number of samples should be accumulated when the number of bits resolved per stage is low. The other is the dither magnitude constraint. The signal range needs to be reduced so that the signal plus dither may not exceed the full-scale range of the MDAC. A signal-dependent dithering scheme can overcome these constraints, which are common in the fixed-magnitude PN dithering [26]. In the signal-dependent dithering, dithers of different magnitudes are selectively injected depending on the signal level so that the signal-to-dither ratio can be minimized, and thereby both constraints can be relieved. When applied to the 1.5 b/stage pipelined ADC, the inter-stage gain error of the standard tri-level MDAC can be measured to be 15 b accurate with a practical number of 2^{26} measurements.

Figure 1.6 shows an example of the pipelined ADC using a tri-level MDAC. In this example, the dither is injected into stage 2 and subtracted digitally from the signal path. The digitized residue of stage 2 is PN-correlated to update the DV_{ref2}. The un-calibrated back-end ADC can be modeled as a linear ADC with a gain error. In the two-capacitor MDAC, the sampled input is amplified by the gain of 2, and bV_{ref} is subtracted depending on the tri-level bit b. The comparator thresholds of the sub-ADC are set to $\pm(1/4)V_{ref}$, and the amplified residue is affected by two major

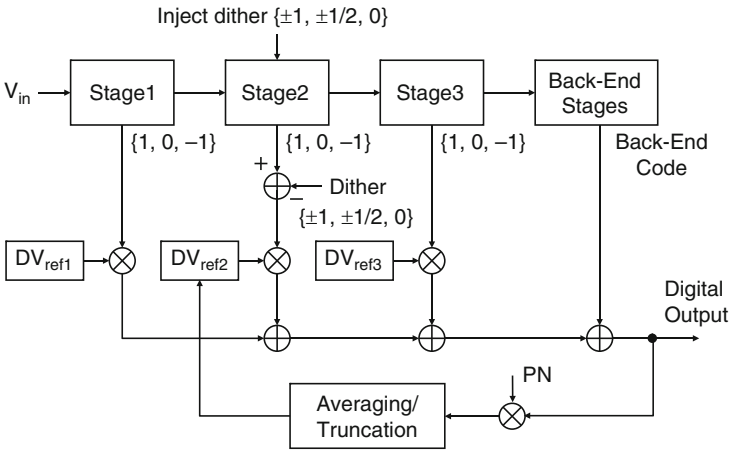


Fig. 1.6 LMS-based update of digital V_{ref}

non-ideal factors, the capacitor mismatch and the finite opamp gain. The digital output is obtained by adding the digital bV_{ref} to the digitized residue, and then divided by 2 so that what is subtracted in the MDAC can be restored digitally. However, the analog bV_{ref} subtracted does not match with the ideal digital bV_{ref}.

1.5.1 Measurement Time and Dither Magnitude Constraints

To measure this non-ideal gain of V_{ref}, a PN-modulated calibration signal V_{CAL}, which is usually a fraction of V_{ref}, is added as a dither into the stage to calibrate. After multiplied by the same PN, it is scaled by V_{ref}/V_{CAL}. The PN-modulated calibration signal is correlated by the same PN sequence and becomes a DC value since $PN^2 = 1$. Therefore, the gain of V_{ref} is obtained by low-pass filtering the digital output. As the bandwidth of the low-pass filter is limited, the noise-like PN-modulated residue remains as a measurement error after low-pass filtered. The measurement error approaches zero if infinitely many samples are averaged. However, as the number of samples is limited in practice, the signal-to-dither ratio should be kept as small as possible in order to minimize the measurement error.

The measurement time constraint results from the tradeoff between the measurement accuracy and the averaging time. Simulations in Fig. 1.7 show that 99% of the measurement errors are smaller than 2^{-10}, which is a 10 b accuracy after averaging 2^{20} samples. Note that four times more samples should be averaged to get one more bit of measurement accuracy. This is true if the PN-modulated residue is treated as a white noise since the standard deviation of the white noise is reduced by the square root of 2 as the number of averaging samples is doubled. Therefore, 2^{30} samples need to be averaged to get the 15 b accuracy, and it takes almost 1 min to complete one measurement if the ADC works at 20 MS/s. The dither magnitude constraint

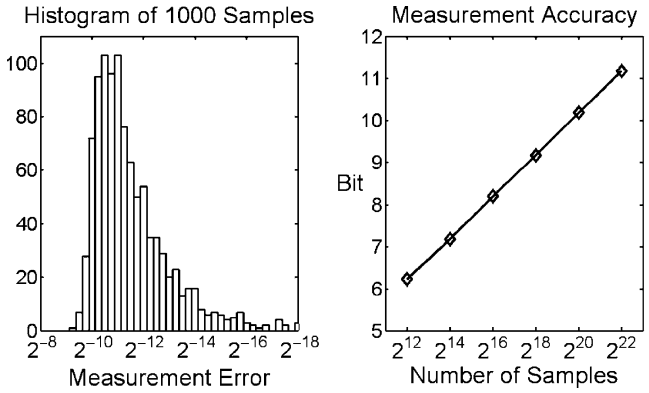


Fig. 1.7 Simulations for correlation accuracy

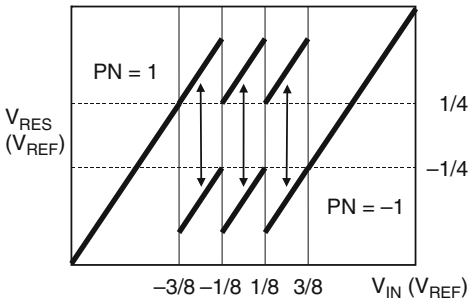


Fig. 1.8 Residue plot for signal-dependent dithering

results from the tradeoff between the dither magnitude and the signal range. The signal range is reduced accordingly to keep the total signal plus dither within the full-scale range, which leads to the reduction in the effective number of bit (ENOB). The signal range reduction is not desirable in a system where the signal-noise ratio (SNR) is dominated by the thermal noise. Switched-capacitor circuits will need capacitors of twice the size to suppress the kT/C noise by 3 dB, thus resulting in a significant area and power penalty. Although a smaller dither makes the signal range larger, it takes much longer to achieve the same accuracy since the signal-to-dither ratio is large. Any solution needs to satisfy both constraints.

1.5.2 Signal-Dependent Dithering Under Two Constraints

Figure 1.8 shows the residue plot of a tri-level MDAC for the full-range signal-dependent dithering. The comparator thresholds in the sub-ADC are shifted from $\pm(1/4)V_{ref}$ to $\pm(3/8)V_{ref}$, and two more comparators are added with thresholds at $\pm(1/8)V_{ref}$ to divide the residue plot into five sub-ranges. A dither of

$-V_{ref}$, $-(1/2)V_{ref}$, 0, $+(1/2)V_{ref}$ or $+V_{ref}$ is injected depending on the PN values and the signal level. No dither is injected when the signal is large for simplicity. Only the dithering of the first stage is sensitive to the input condition while the later stages are not. Therefore, the delay in the measurement for not dithering when the signal is large is insignificant. The signal plus dither between $\pm(3/8)V_{ref}$ is in effect a large fixed-magnitude dither of $(1/2)V_{ref}$ with a small signal within the range of $\pm(1/4)V_{ref}$.

Signal-dependent dithering still offers a substantial saving in the measurement time with low circuit complexity unless the signal stays at a high level all the time. The signal-to-dither ratio of V_{in}/V_{CAL} is reduced to 1/2, and 99% of the measurement errors are smaller than 2^{-14} when only 2^{26} samples are averaged. If referred to the input after divided by 2, it corresponds to 15 b accuracy.

The standard tri-level MDAC is modified for the signal-dependent dithering by adding two more comparators and splitting one of the capacitors into two as shown in Fig. 1.9. Dithers are injected by controlling the switches according to the comparator outputs and PN values. Both C1 and C2 are switched between $-V_{ref}$ and 0 for the signal range from $-(3/8)$ to $-(1/8)V_{ref}$, and between 0 and $+V_{ref}$ for the signal range from $+(1/8)$ to $+(3/8)V_{ref}$ if PN is 1 and -1, respectively. When the signal lies in the middle range, C1 and C2 are alternately switched to $-V_{ref}$ if PN $= 1$ and $+V_{ref}$ if PN $= -1$ to inject a dither of $(1/2)V_{ref}$ equally through two capacitors. The mismatch between the two split capacitors contributes to noise after randomized and spread over the Nyquist band. It needs to be subtracted digitally. The proposed tri-level MDAC has the following features. (1) Large dithers are used without sacrificing the signal range. (2) The signal de-correlation time is greatly shortened due to the low signal-to-dither ratio. (3) No additional capacitor is used for dithering, and the analog performance is not affected. (4) Switch logic doesn't delay opamp settling.

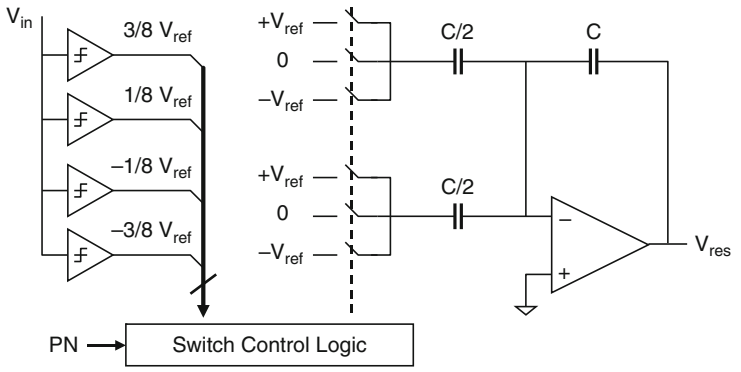


Fig. 1.9 MDAC and comparators for signal-dependent dithering

1.5.3 Linearity Improvement

The end result of the DAC/gain error calibration is dramatic in the measured INL and FFT. A prototype fabricated in 0.18 μ CMOS occupies $2.3 \times 1.7\,\text{mm}^2$. The digital logic occupies $0.6\,\text{mm}^2$. The sampling capacitors in the S/H and stages 1–4 are set to 2 pF, and the kT/C noise limits the SNR to be −76 dB with 2 V_{pp} full-scale range. Stages 5–14 are scaled down by half to save the chip power and area.

Figure 1.10 shows the measured INL at a 15 b level before and after calibration sampled at 20 MS/s. The INL error jumps significantly at the comparator threshold points before calibration. The largest INL jump is at the first stage comparator thresholds. After the first six stages are calibrated, the INL errors are greatly reduced and improved from 25 LSB to 1.3 LSB. The FFT of a 14.5 MHz input sampled at 20 MHz is also shown. The ADC linearity is improved to 15 b while the SNDR is mainly limited by the kT/C noise.

It takes 45 s to calibrate the first six stages with a full-scale sinusoidal input at 20 MS/s. The calibration time is reduced to 38 s if the input is random within the full-scale input range since the sinusoidal signal gives less number of samples at low signal levels. This calibration time difference is not significant as mentioned before since only the calibration time of the first stage is sensitive to the input level. The advantage of the signal-dependent dithering is obvious. In the previous work of a 1.5 b/stage pipelined ADC [20], which loses 25% of the signal range and averages 8×2^{28} samples per stage, it took 8.95 min to calibrate five stages at 20 MS/s while achieving less calibration accuracy than this example. The calibration time can be further saved by gradually scaling down the measurement accuracy. For example, by scaling 0.5 b accuracy per two stages, it can be reduced to 24 s with a random input. Higher sampling rate is also effective in shortening the calibration time. The prototype consumes 285 mW @1.8 V. Performance of high-resolution ADCs is severely degraded without input and clock buffers. Consuming the same power, the same ADC in different versions works at 60 MS/s with 15 b linearity.

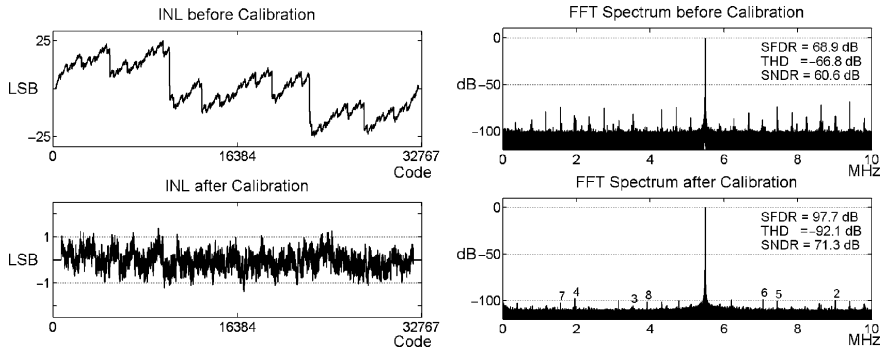


Fig. 1.10 INL and FFT before and after calibration

1.5.4 Opamp Non-linearity Calibration

The circuit complexity of the residue amplifier grows due to the high gain and wide bandwidth requirements. In opamps with low supply voltages, the non-linearity is a dominant factor limiting the residue accuracy. The opamp non-linearity effect appears in the residue output of the 3-b tri-level MDAC example as shown in Fig. 1.11. It can create discontinuities in the transfer function like missing codes. The discontinuity can be removed by calibration like DAC and gain errors if the digital steps at the comparator thresholds can be measured.

However, the residual nonlinearity still remains. Unlike the DAC and gain calibration, which calibrates errors only at major comparator thresholds, the opamp non-linearity calibration is very close to the code mapping for the entire transfer function that requires a long training or measurement cycle. A more realistic solution is to approximate the opamp non-linearity with a high-order polynomial as shown in Fig. 1.12. In foreground measurements, it is easy to try several input levels to map the opamp transfer function, but in background measurements, it is difficult to use large dithers to measure the transfer function.

Three compromises have been proposed to date for the background opamp non-linearity measurement. All of them assume that the opamp is weakly non-linear so

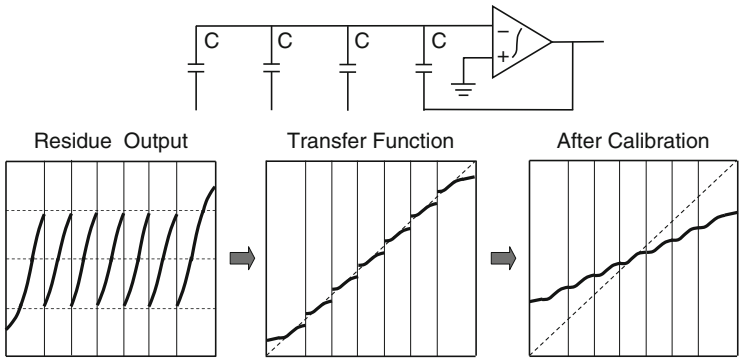


Fig. 1.11 Opamp non-linearity effect on ADC transfer curve

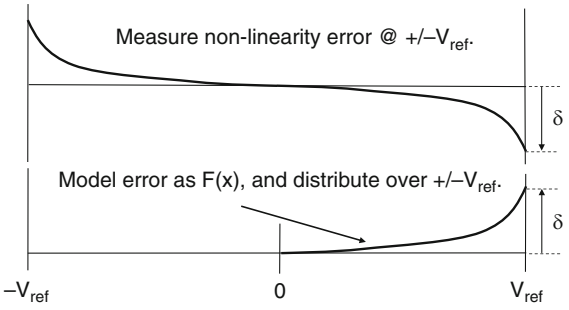


Fig. 1.12 Curve fitting of opamp non-linearity error

that the third-order distortion can be modeled as a dominant term [27–29]. Heavily non-linear cases may need more complicated higher-order curve-fitting or calibration. One is to use a code density histogram to measure the gain errors, and distribute them using a look-up table over the range [27]. It assumes that the random signal covers the measurement range to give the sufficient code density for all codes. The others are to use multi-level PN dithers to estimate the third-harmonic distortion term [28, 29]. However, the non-linearity calibration has yet to achieve such a high linearity on par with what is feasible with just the DAC and gain calibration. It has been proved to exhibit 12–13 b resolution, which is sufficient to show the proper ADC operation using non-linear opamps. ADC designers may need to go this extra distance to ensure that ADCs they design work in the low-voltage environment. The following CT $\Delta\Sigma$ approach may offer an alternative route to reach the same goal.

1.6 Noise Leakage Calibration in CT Cascaded $\Delta\Sigma$ Modulator

While the pipelined ADC is being calibrated, the CT $\Delta\Sigma$ modulator has also been updated with scaled digital technologies. Its advantage is that the CT filter performs anti-aliasing, and the quantized feedback is far less sensitive to the non-linearity as they are reduced by the filter gain. The input sampling jitter is not an issue since the sampling is done after the filter, but the jitter is critical in the feedback DAC path. Due to the pulse width jitter problem, either the SC DAC or multi-bit DAC have been used. To achieve high resolution with a low over-sampling ratio of 6–8, the modulator order should be higher than 4, and 3–4 b DACs have been used. Two high-order architectures can be considered. One is the single-loop modulator, and the other is the cascaded one.

What cascaded is to single-loop for $\Delta\Sigma$ modulators is what pipelined is to flash for Nyquist ADCs. The stability of the higher-order single-loop modulator has been an issue. Cascading low-order stages can achieve wide bandwidth with low OSR without the stability concern. In cascaded $\Delta\Sigma$ modulators, a digital noise cancellation filter (NCF) is used to remove the quantization noise from the earlier stages and also to shape that of the last stage. The noise leakage in the DT cascaded modulator results from the capacitor mismatch and finite opamp gain, but in the CT cascaded modulator, there are several factors that cause incomplete noise cancellation. The noise leakage is the same problem as the reference mismatch in the pipeline discussed earlier.

One factor to affect the noise leak is the accuracy of the CT-to-DT transform. The exact transform varies depending on the actual shape of the DAC pulse, and may involve complicated calculations [30]. An earlier work with a 4 b quantizer uses a modified bilinear CT-to-DT transform to approximate the in-band frequency response [31], but the limited amount of noise suppression may not be sufficient if low-resolution quantizers are used. The other is the variation of the RC or C/Gm time-constant of the loop filter over process, voltage, and temperature. Previous works adjust the digital NCF [31] or variable resistors [32] to minimize the in-band

digital output noise, but the input of the modulator should be forced to be zero while calibrating. A simplified CT-to-DT transform is first derived to find the exact NCF, and the filter time-constant is calibrated in background based on the LMS adaptation interrupting the normal operation.

1.6.1 CT-to-DT Transform

The CT-to-DT transform is to find the DT counterpart of a CT filter so that the CT DAC output waveform sampled by the quantizer can match that of the DT DAC output [30]. The transform is affected by the CT filter types and DAC pulse shapes, and difficult to derive analytically. A parameter-based approach is devised to find the exact DT counterparts of such CT integrators as $1/s$, $1/s^2$, $1/s^3$, etc. The same approach can be generalized to derive other transforms such as for CT resonators. Shown in Fig. 1.13 is a DAC pulse between $t = 0 - T_S$ that passes through a series of integrators with a time-constant T_S. The integrator outputs when $t \geq T_S$ can be expressed as simple polynomials of $(t - T_S)/T_S$, as depicted with solid lines. The coefficients of the polynomials are set by the parameters of a, b, and c. They are the outputs of the first, second, and third integrators when sampled at $t = T_S$, respectively.

The time-domain polynomials are then converted into DT functions using z-transform with a sampling period of T_S so that the DT functions can have the

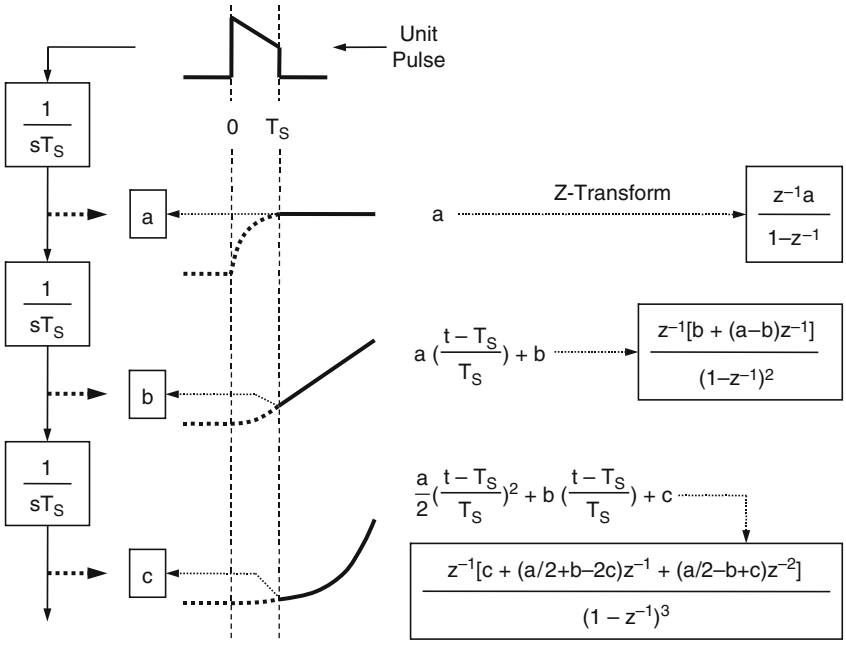


Fig. 1.13 Principle of parameter-based CT-to-DT transform

same pulse response as the CT filters with a sampling rate of $1/T_S$. This can be developed into a look-up table approach including any different types of CT filters. Designers can use such a table to find the DT functions and get the parameters at $t = T_S$ by simulations using Matlab or SPICE with real DACs followed by the filtering functions. In practice, the DAC output is delayed to avoid the comparator meta-stability. A delayed pulse can be divided into two pulses between $0 - T_S$ and $T_S - 2T_S$. The later one can be handled as a pulse between $0 - T_S$ delayed by one full cycle.

1.6.2 Calibrated Cascaded ΔΣ Modulator

To calibrate the filter time-constant variation accurately without interruption, the self-tuning technique used for the single-stage modulator [33] can be modified for the cascaded modulator. The calibration block is shown at the top of Fig. 1.14. A single binary tone at f_{tone} is injected into the first-stage quantizer input. It is

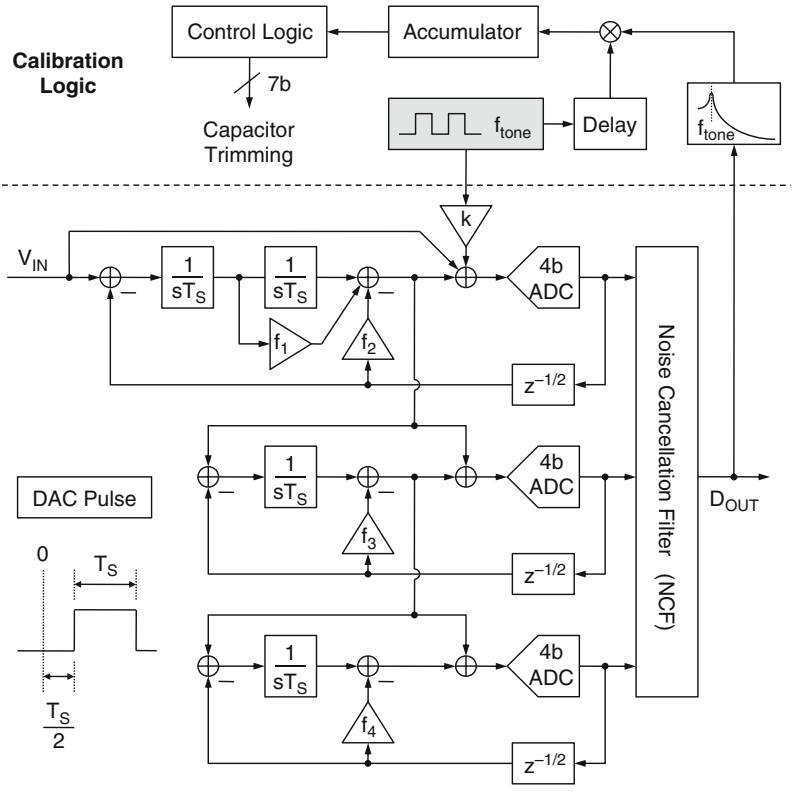


Fig. 1.14 2–1–1 Cascaded CT ΔΣ modulator and calibration block

therefore considered as a part of the quantization noise, and should be cancelled by the digital NCF. If the analog filter and the digital NCF are not matched, a residual tone appears in the digital output.

Since the polarity of the residual tone is the same as that of the time-constant error, the time constant can be tuned using the zero-forcing adaptive LMS feedback. The residual tone polarity is detected by correlating the digital output with the same injected pulse. An IIR filter amplifies the residual tone f_{tone} and suppresses its harmonics before correlation to shorten the correlation time and to avoid the harmonic mixing. Note that the f_{tone} can be either inside or outside the signal band since the noise cancellation works for all frequencies.

The lower part of Fig. 1.14 shows the block diagram of a 2–1–1 cascaded CT $\Delta\Sigma$ modulator example with all the NTF zeros placed at DC. It uses half-cycle delayed 4 b current DACs to reduce the effects of the clock jitter sensitivity and comparator meta-stability. This extra half-clock delay is compensated for by the quantizer feedbacks of f_2–f_4. In every stage, a feed-forward path is added from the input to the quantizer input so that the loop filter output can be directly connected to the next stage without using extra DACs. The coefficients are chosen for stability and performance. The NCF is then derived to cancel the quantization noises of the first and second stages. The CT-to-DT transform of every possible path from the DAC output to the quantizer input should be considered. This is different from finding the NCF of a DT modulator. In fact, the system can be configured with zeros placed at any frequencies using resonators, and their CT-to-DT transforms can be derived.

The simulated FFT spectrum shown in Fig. 1.15 exhibits a fourth-order noise shaping of 80 dB/decade. The benefit of using an exact NCF is evident. While the previous 2–2 cascaded design has a simulated signal-to-quantization-noise ratio (SQNR) of 79 dB at 8 × OSR with optimally placed NTF zeros [31], this example achieves a similar SQNR of 77 dB at this low OSR with all the NTF zeros at DC.

A prototype in 0.18 μ CMOS sampling at 360 MHz is dithered with a pulse f_{tone} of $\pm 1/4$ LSB at 18 MHz. The binary-weighted capacitor banks in the Gm-C filters are trimmed with a 1.1% step. Figure 1.16a shows the measured residual tone magnitude with different capacitor tuning codes. The residual tone magnitude is

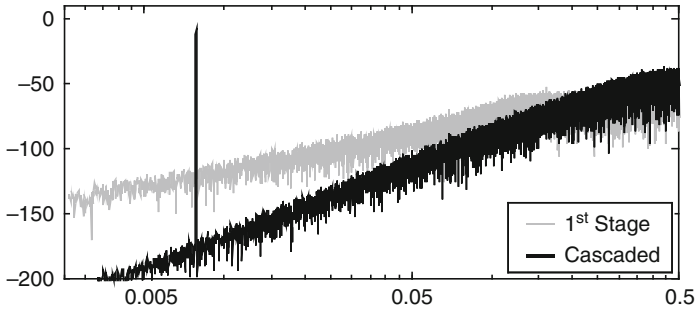


Fig. 1.15 Simulated FFT spectrum of 2–1–1 CT $\Delta\Sigma$ modulator

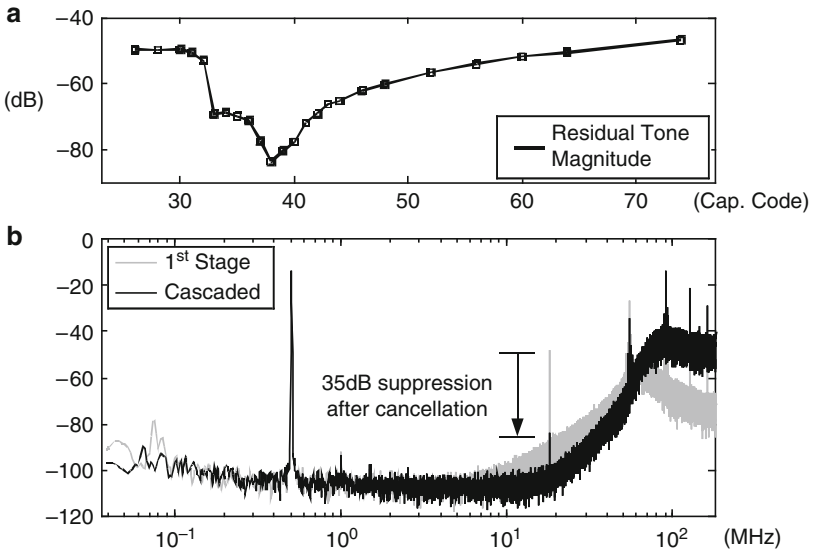


Fig. 1.16 **(a)** Residual tone vs. tuning. **(b)** FFT spectrum before and after

detected with an accuracy of better than 1.1%, and is suppressed to −84 dBFS after calibration. Figure 1.16b shows the measured FFT spectrums of the modulator output and the first-stage output. The f_{tone} at 18 MHz, which is second-order shaped in the first stage, is suppressed by 35 dB after the adaptive noise cancellation. The high-frequency droops in the FFTs result from the high-frequency poles, and the low-frequency noise is dominated by the thermal noise.

1.7 Conclusions

Digital techniques are bound to affect how data converters are designed. Many high-speed ADC architectures will emerge, and in the high-resolution arena, the calibrated pipelined ADC and the CT $\Delta\Sigma$ modulator would compete. The past history tells us that the over-sampling feedback approach would overtake the Nyquist-rate pipelined approach. In fact, the CT $\Delta\Sigma$ approach is more desirable in most SOCs as it includes the anti-aliasing function. However, with low-voltage scaled CMOS, even the CT $\Delta\Sigma$ modulator with low single-digit over-sampling ratios experiences the same difficulties as the pipelined ADC. Therefore, digital assisting will be at the center of most future ADC designs. In particular, the LMS-based adaptive zero-forcing feedback ensures that digital assisting will work in a more robust way.

References

1. Z. Boyacigiller, B. Weir, and P. Bradshaw, "An error-correcting 14 b/20 μs CMOS A/D converter," *ISSCC Dig. Tech. Papers*, pp. 62–63, Feb. 1981.
2. J. Domogalla, "Combination of analog to digital converter and error correction circuit," *US Patent*, 4451821, May 1984.
3. H. S. Lee, D. A. Hodges, and P. R. Gray, "A self-calibrating 15 bit CMOS A/D converter," *IEEE J. Solid-State Circuits*, vol. SC-19, pp. 813–819, Dec. 1984.
4. S. H. Lewis and P. R. Gray, "A pipelined 5-Msample/s 9-bit analog-to digital converter," *IEEE J. Solid-State Circuits*, vol. SC-22, pp. 954–961, Dec. 1987.
5. B. S. Song, M. F. Tompsett, and K. R. Lakshmikumar, "A 12-bit 1 Msample/s capacitor error-averaging pipelined A/D converter," *IEEE J. Solid-State Circuits*, vol. SC-23, pp. 1324–1333, Dec. 1988.
6. B. S. Song, S. H. Lee, and M. F. Tompsett, "A 10-b 15-MHz CMOS recycling two-step A/D converter," *IEEE J. Solid-State Circuits*, vol. SC-25, pp. 1328–1338, Dec. 1990.
7. Y. Lin, B. Kim, and P. Gray, "A 13-b 2.5-MHz self-calibrated pipelined A/D converter in 3 μ CMOS," *IEEE J. Solid-State Circuits*, vol. SC-26, pp. 628–636, Apr. 1991.
8. MAX1200, "+5 V Single-Supply, 1MS/s, 16-Bit Self-Calibrating ADC," *Maxim*, 1998.
9. S. H. Lee and B. S. Song, "Digital-domain calibration of multistep analog-to-digital converters," *IEEE J. Solid-State Circuits*, vol. SC-27, pp. 1679–1688, Dec. 1992.
10. A. N. Karanicolas, H. S. Lee, and K. L. Bacrania, "A 15-b 1-Msample/s digitally self-calibrated pipeline ADC," *IEEE J. Solid-State Circuits*, vol. SC-28, pp. 1207–1215, Dec. 1993.
11. T. H. Shu, B. S. Song, and K. Bacrania, "A 13-b 10-Msample/sec ADC digitally calibrated with real-time oversampling calibrator," *IEEE J. Solid-State Circuits*, vol. SC-30, pp. 443–452, Apr. 1995.
12. U. K. Moon and B. S. Song, "Background digital calibration techniques for pipelined ADCs," *IEEE Trans. Circuits Syst. II*, vol. 44, pp. 102–109, Feb. 1997.
13. S. U. Kwak, B. S. Song, and K. Bacrania, "A 15-b, 5-Msample/s low-spurious CMOS ADC," *IEEE J. Solid-State Circuits*, vol. SC-32, pp. 1866–1875, Dec. 1997.
14. J. Ingimo and B. Wooley, "A continuously calibrated 12-b, 10-MS/s, 3.3-V A/D converter," *IEEE J. Solid-State Circuits*, vol. SC-33, pp. 1920–1930, Dec. 1998.
15. O. E. Erdogan, P. J. Hurst, and S. H. Lewis, "A 12-b digital-background-calibrated algorithmic ADC with −90-dB THD," *IEEE J. Solid-State Circuits*, vol. SC-34, pp. 1812–1820, Dec. 1999.
16. R. C. Dixon, "Spread Spectrum Systems," *New York: Wiley*, 1976.
17. R. Jewett, K. Poulton, K. C. Hsieh, and J. Doernberg, "A 12b 128 Msample/s ADC with 0.05LSB DNL," *ISSCC Dig. Tech. Papers*, pp. 138–139, Feb. 1997.
18. J. Ming and S. H. Lewis, "An 8-bit 80-Msample/s pipelined analog-to-digital converter with background calibration," *IEEE J. Solid-State Circuits*, vol. 36, pp. 1489–1497, Oct. 2001.
19. E. Siragursa and I. Galton, "A digitally enhanced 1.8-V 15-bit 40-MSample/s CMOS pipelined ADC," *IEEE J. Solid-State Circuits*, vol. 39, pp. 2126–2138, Dec. 2004.
20. H. C. Liu, Z. M. Lee, and J. T. Wu, "A 15-b 40-MS/s CMOS pipelined analog-to-digital converter with digital background calibration," *IEEE J. Solid-State Circuits*, vol. SC-40, pp. 1047–1056, May 2005.
21. B. Widrow, J. McCool, and M. Ball, "The complex LMS algorithm," *IEEE Proc*., vol. 59, p. 719, Apr. 1971.
22. C. H. Heng, M. Gupta, S. H. Lee, D. Kang, and B. S. Song, "A CMOS TV tuner/demodulator IC with digital image rejection," *IEEE J. Solid-State Circuits*, vol. SC-40, pp. 2525–2535, Dec. 2005.
23. S. Lerstaveesin and B. S. Song, "A complex image rejection circuit using sign detection only," *IEEE J. Solid-State Circuits*, vol. SC-41, pp. 2693–2702, Dec. 2006.
24. M. Gupta and B. S. Song, "A 1.8 GHz spur-cancelled fractional-N frequency synthesizer with LMS-based DAC gain calibration," *IEEE J. Solid-State Circuits*, vol. SC-41, pp. 2842–2851, Dec. 2006.

25. S. T. Ryu, S. Ray, B. S. Song, G. H. Cho, and K. Bacrania, "A 14-b linear capacitor self-trimming pipelined ADC," *IEEE J. Solid-State Circuits*, vol. SC-39, pp. 2046–2051, Nov. 2004.
26. Y. S. Shu and B. S. Song, "A 15 b-linear, 20 MS/s, 1.5 b/stage pipelined ADC digitally calibrated with signal-dependent dithering," *IEEE J. Solid-State Circuits*, vol. SC-43, pp. 342–350, Feb. 2008.
27. B. Murman and B. Boser, "A 12 b 75 MS/s pipelined ADC using open-loop residue amplification," *IEEE J. Solid-State Circuits*, vol. SC-39, pp. 2040–2050, Dec. 2003.
28. J. P. Keane, P. J. Hurst and S. H. Lewis "Background interstage gain calibration technique for pipelined ADCs," *IEEE Trans. Circuits Syst. I*, vol. 52, pp. 32–43, Jan. 2005.
29. A. Panigada and I. Galton, "A 130 mW 100 MS/s pipelined ADC with 69 dB SNDR enabled by digital harmonic distortion correction," *ISSCC Dig. Tech. Papers*, pp. 162–163, Feb. 2009.
30. O. Oliaei, "Design of continuous-time sigma-delta modulators with arbitrary feedback waveform," *IEEE Trans. Circuits Syst. II*, vol. 50, no. 8, pp. 437–444, Aug. 2003.
31. L.J. Breems, R. Rutten, and G. Wetzker, "A cascaded continuous-time $\Sigma\Delta$ modulator with 67-dB dynamic range in 10-MHz bandwidth," *IEEE J. Solid-State Circuits*, vol. 39, no. 12, pp. 2152–2160, Feb. 2004.
32. L.J. Breems, R. Rutten, R.H.M. van Veldhoven, and G. van der Weide, "A 56 mW continuous-time quadrature cascaded $\Sigma\Delta$ Modulator with 77 dB DR in a near zero-IF 20 MHz band," *IEEE J. Solid-State Circuits*, vol. 42, no. 12, pp. 2696–2705, Dec. 2007.
33. Y. S. Shu, B. S. Song, and K. Bacrania, "A 65 nm CMOS CT $\Delta\Sigma$ modulator with 81 dB DR and 8 MHz BW auto-tuned by pulse injection," *ISSCC Dig. Tech. Papers*, pp. 500–501, Feb. 2008.

Chapter 2
Pipelined ADC Digital Calibration Techniques and Tradeoffs

Imran Ahmed

Abstract In this paper an overview of state of the art techniques to measure and correct non-idealities in a pipelined ADC is given. The paper discusses the motivations for digital calibration, and subsequently details state of the art calibration approaches. System tradeoffs of commonly used calibration techniques are analyzed. A discussion of how digital calibration can be used to enable the next generation of very low power 'smart-ADCs' is also given.

2.1 Introduction

The pipelined topology is a popular option for ADCs which require resolutions on the order of 8 to 14 b and sampling rates between a few MS/s to hundreds of MS/s. The popularity of the topology can be attributed to its relatively simple and repetitive core structure, as well as a significant reduction in the number of comparators required to achieve a fixed resolution when compared to other Nyquist-rate data converters such as Flash, folding + interpolating, etc. Pipelined ADCs are used in a variety of applications such as: mobile systems, CCD imaging, ultrasonic medical imaging, digital receivers, base stations, digital video (e.g. HDTV), xDSL, cable modems, and fast Ethernet. With the use of pipelined ADCs in many consumer products, research in improving the performance of pipelined ADCs has attracted much attention over the past decade, where the most popular areas of research have been: linearity enhancement, and power reduction.

Linearity enhancement has been an active area of research as with deep submicron technology low intrinsic gain from MOSFETs, low supply voltages, and device mismatch have made achieving very linear data converters (i.e. >10-b linear) challenging using conventional pipelined ADC design techniques. Low power consumption in pipelined ADCs is motivated by the fact that for mobile systems which use pipelined ADCs, low power consumption enables increased battery life

I. Ahmed (✉)
Kapik Integration, 192 Spadina Ave., Suite 406, Toronto, Ontario, M5T 2C2, Canada
e-mail: imran.ahmed@utoronto.ca

A.H.M. van Roermund et al. (eds.), *Analog Circuit Design: Smart Data Converters, Filters on Chip, Multimode Transmitters*, DOI 10.1007/978-90-481-3083-2_2,
© Springer Science+Business Media B.V. 2010

and thus increased user productivity. In wired systems where many ADCs can be integrated on-chip in parallel, power savings enable cheaper packaging.

In this paper digital techniques which enable enhanced linearity in pipelined ADCs, thus relaxed design constraints for analog circuits and hence lower power consumption, will be discussed. In Sect. 2.2, a review of the pipelined ADC and error sources which require calibration will be given. In Sect. 2.3, digital calibration including foreground and background techniques will be examined with the associated tradeoffs of each approach detailed. In Sect. 2.4 techniques to enable rapid background digital calibration, and thus address many of the tradeoffs of background calibration noted in Sect. 2.3 will be discussed. In Sect. 2.5 a topology to exploit digital calibration so as to enable very low power consumption in the next generation of 'smart ADCs' will be given. Section 2.6 concludes the paper.

2.2 Review of Error Sources in Pipelined ADCs

In Fig. 2.1 the topology of a typical pipelined stage (4-b example shown, including 1-b redundancy to relax sub-ADC requirements) is illustrated. In Fig. 2.2 an example circuit implementation of the pipelined stage topology is displayed. Figure 2.3 illustrates the input/output plot (residue transfer curve) of the pipelined stage when no errors are present. In the following sub-sections the impact of the dominant and most commonly corrected errors: Gain, and DAC errors, will be analyzed.

2.2.1 Gain Errors

Consider the practical situation where due to mismatch between the sampling capacitors C_0 to C_{15} and the feedback capacitor C_f and also due to low DC gain from

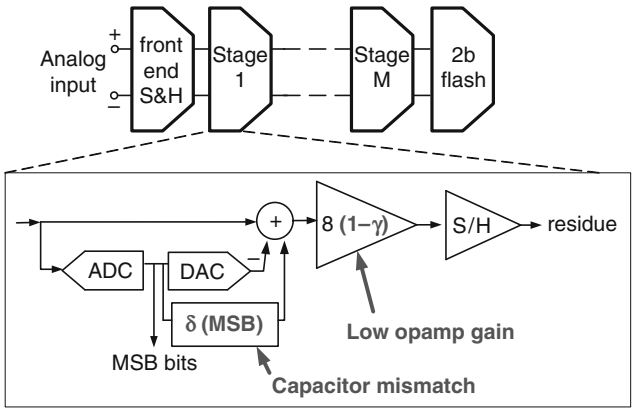


Fig. 2.1 Pipeline topology, first stage shown in detail including error sources

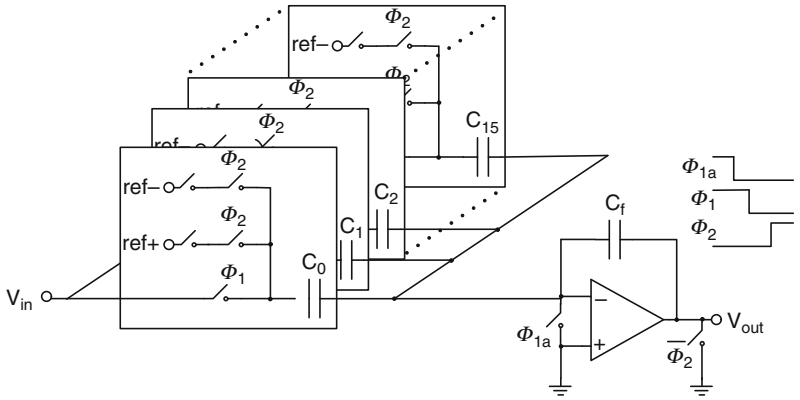


Fig. 2.2 Example implementation of 4-b MDAC

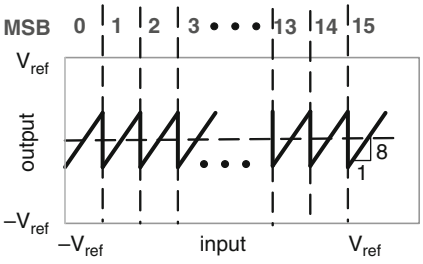


Fig. 2.3 Ideal residue transfer curve of 4-b pipeline stage

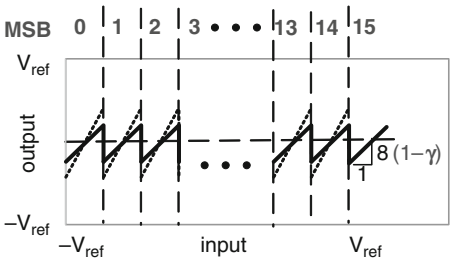


Fig. 2.4 Residue transfer curve showing impact of gain errors

the opamp in Fig. 2.2, the ideal gain of 8× of a 4-b pipeline stage is modified by $(1-\gamma)$. As shown in Fig. 2.4 the modified stage gain results in a fixed number of missing codes at every MSB transition (i.e. constant DNL errors or constant jumps in INL at every transition of the bits resolved by the first stage).

Common analog techniques to reduce gain errors below the LSB level include: using very large capacitors to sufficiently minimize capacitor mismatch, and/or using gain boosting [1], multi-stage opamp [2] techniques, or using long channel lengths for key transistors to achieve very large DC opamp gains. Using large capacitors, opamp gain enhancing techniques, and long channel lengths however

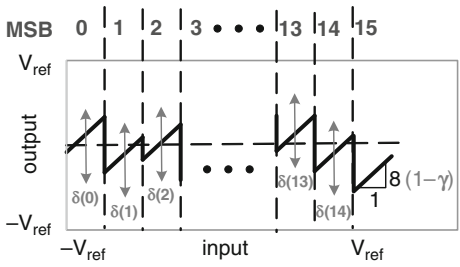


Fig. 2.5 Residue transfer curve showing impact of DAC and gain error

come at the penalty of increased power consumption. Furthermore due to technology limitations capacitor mismatch and opamp gain cannot be arbitrarily improved using analog techniques.

2.2.2 DAC Errors

Capacitor mismatch between *each* of the sampling capacitors C_0 to C_{15} in Fig. 2.2 results in errors in the pipeline stage's DAC which are a function of each MSB bit resolved. As shown in Fig. 2.5, DAC errors result in *each* linear segment of the residue transfer curve being shifted up or down by different static random values $\delta(i)$. Hence DAC errors result in a different number of missing codes at every MSB transition, yielding substantial harmonic distortion.

The common analog technique to minimize DAC errors is to use large capacitors. However as discussed in Sect. 2.2.1, this comes at the cost of increased power, and due to technology limitations capacitor mismatch cannot be made arbitrarily small.

2.3 Digital Calibration Techniques

As the outputs of ADCs are ultimately digital, rather than correcting the non-idealities of pipelined ADCs in the analog domain, the non-idealities can be corrected by manipulating the digital output of the ADC as shown in Fig. 2.6.

By correcting the ADC errors in the digital domain, analog design requirements can be relaxed (e.g. smaller capacitors can be used, lower DC gain in opamps, minimum size devices). Since analog power consumption is generally much larger than digital power consumption in deep sub-micron processes, the trade-off of correcting the non-ideality in the digital domain generally results in an overall reduction of power. Furthermore, as technology scaling tends to favor digital circuitry over analog circuitry, it becomes even more desirable in newer technology nodes to trade analog circuitry with digital circuitry.

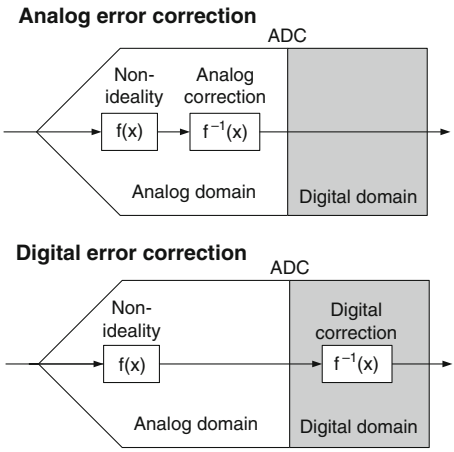


Fig. 2.6 Analog versus digital error calibration

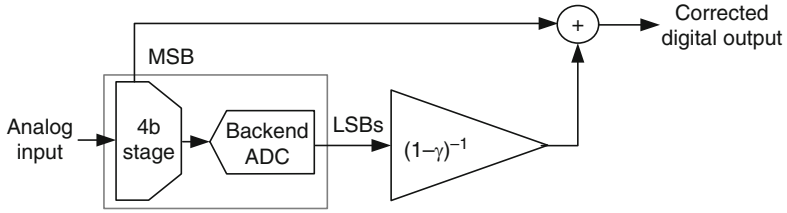


Fig. 2.7 Gain error correction of first pipeline stage

2.3.1 Digital Gain Error Calibration

Gain error can be digitally corrected by scaling the digital output of the backend ADC by the inverse of the gain error factor $(1-\gamma)$. Figure 2.7 illustrates an example of an architecture which compensates for the effect of the non-ideality γ in the first pipeline stage, assuming the value of γ is already known. The entire pipeline can be calibrated by starting calibration with the last pipeline stage and recursively using the same technique to calibrate earlier pipeline stages [3].

2.3.2 DAC Gain Error Calibration

From Sect. 2.2.2 it was shown that capacitor mismatch in the DAC results in unique missing codes at every MSB transition, thus a separate corrective term for *each* MSB transition is required, significantly increasing the complexity of the correction scheme over gain-only correction techniques. For example, with a 3 + 1-b pipeline stage, 15 correction parameters for 16 unique DAC outputs are required to be estimated, whereas a gain-only correction scheme has only one parameter to estimate.

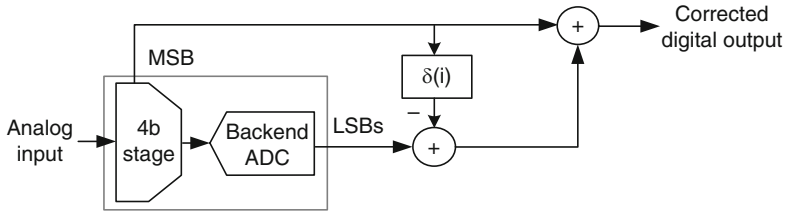


Fig. 2.8 Correction of gain and DAC errors in first pipeline stage

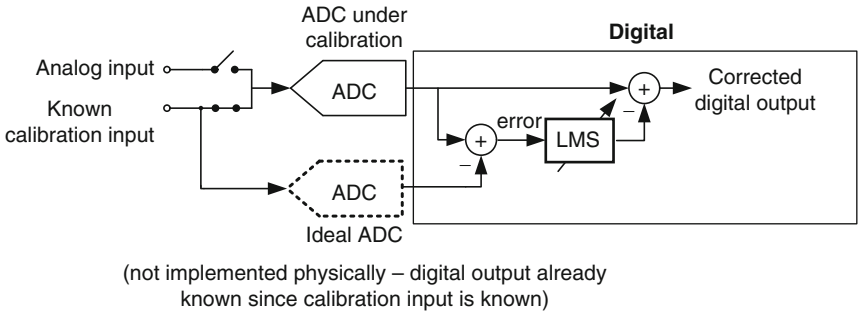


Fig. 2.9 Principle of foreground calibration

If the amount of DAC error is known, DAC errors can be corrected by simply shifting the digital ADC output as a function of the MSB by the negative amount of the missing codes produced by the DAC errors, as shown in Fig. 2.8.

Comparing Figs. 2.4 and 2.5, it is noted that missing codes produced by gain errors look the same as missing codes produced by DAC errors where the DAC error is constant at every MSB transition. Thus in a DAC calibration scheme (where the missing codes are corrected as a function of *each* MSB), the gain errors are also corrected in addition to DAC errors.

2.3.3 Foreground Calibration Techniques

Sections 2.3.1 and 2.3.2 discussed how gain and DAC errors can be corrected when the amount of error is already known in advance. In reality however the error is unknown to the designer before fabrication. Furthermore the magnitude of each error source varies from chip to chip due to process variation. Thus a scheme to adaptively measure the unknown and unique error sources in an ADC needs to be implemented. In Fig. 2.9, a foreground calibration scheme is shown.

Foreground calibration estimates the unknown errors sources by interrupting normal ADC operation and applying a known input sequence to the ADC. By comparing the output of the ADC to the expected ADC output under ideal conditions

(i.e. no non-idealities) the impact of each error source can be measured and corrected. Examples of foreground calibration in publications can be found in [4] and [5].

The advantage of a foreground scheme is that calibration can be achieved within a small number of clock cycles, since the error signal labeled in Fig. 2.9 is highly correlated with the error sources causing the missing codes. The disadvantage of foreground calibration is that the ADC is required to be taken offline every time calibration is performed, which in some applications may not be possible.

2.3.4 Background Calibration

Background calibration continuously measures and corrects the effect of non-idealities in a pipeline stage, thus has the significant advantage that the ADC is not required to be taken offline to perform calibration. As such the vast majority of calibration techniques published are focused on background techniques. Some example publications of ADCs with background calibration can be found in [6–22].

Several topologies have been proposed recently to implement background calibration, where the vast majority of the schemes use a statistics based approach. In a statistical scheme, the input of the pipeline stage under calibration is combined with a known pseudo-random sequence, where by correlating the digital output of the ADC with the known pseudo random sequence, the impact of missing codes can be determined. To avoid significantly altering the ADC output spectrum, the pseudo-noise sequence is typically made very long to avoid correlations with the analog input, as well as small in amplitude so that the injected pseudo-random sequence which appears as an additional white noise source at the output only consumes a small portion of the dynamic range. Figure 2.10 shows the basic principle of statistics based background calibration.

With statistics based background calibration schemes however, since the digital output of the ADC is highly correlated with the analog input and weakly correlated with the pseudo-random sequence, a large number of clock cycles are required to accurately extract the pseudo-random sequence from the digitized analog input in the ADC output. For example, in [16] $\sim 10^7$ cycles were required to achieve 13-b

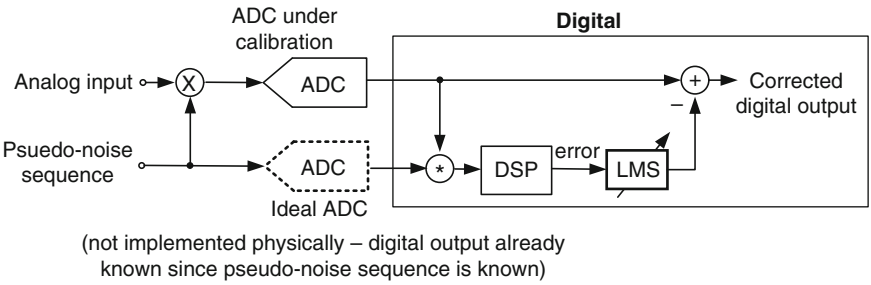


Fig. 2.10 Principle of statistical based background calibration

linearity, and in [17] $\sim 10^8$ clock cycles were required to achieve >14 b linearity. In [18] it was shown empirically that statistical techniques required on the order of 2^{2N} clock cycles to calibrate gain errors only. For 11-b linearity ~4 million clock cycles are required to only correct gain errors using statistics-based background calibration. Thus, while background schemes are popular as they enable continuous ADC operation, the calibration time of background approaches is very lengthy.

In an industrial environment where ICs are mass produced, ICs are tested for functionality by automated testers. In ADCs which use background-statistical techniques to achieve calibration, long calibration times can lead to excessive test times thus limiting IC production throughput and hence revenue. For example, with 4 million calibration cycles, even with a reasonably high sampling rate of 40 MS/s, 1/10th of a second would be required at minimum to test each ADC. For higher resolution and/or lower speed ADCs the test time can be much higher [18]. In the interest of larger production throughput it is highly desirable to reduce calibration time.

2.4 Rapid Background Calibration Techniques

From the discussion of Sect. 2.3, background calibration was shown to be more desirable than foreground calibration as the ADC can continuously operate. However lengthy calibration times of statistical background techniques result in a large penalty in long testing times. In this section, techniques to significantly reduce background calibration times, and thus make background calibration more attractive for industrial products will be discussed in detail.

2.4.1 Slow but Accurate Parallel ADC

The brute force approach to reduce calibration time is to digitize the analog input with two ADCs: a high speed ADC which suffers from gain and DAC errors, and a low speed ADC which is very accurate, as shown in Fig. 2.11 (e.g., [19]).

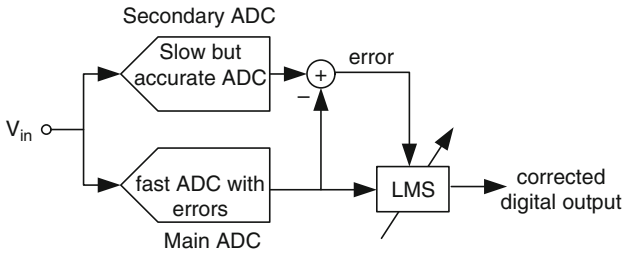


Fig. 2.11 Rapid calibration using secondary, slow but accurate ADC

With the topology of Fig. 2.11, the output of the slow but accurate ADC can be used as an ideal reference to compare against the output of fast ADC (down-sampled output). Since the error signal in Fig. 2.11 is highly correlated with the error sources, calibration can be achieved by only looking at a small number of samples of the error signal. Thus with the brute force method, the digital circuitry of the background scheme becomes simpler, but at the cost of increased analog complexity.

The main drawback of the approach of Fig. 2.11 is that in addition to designing the main path high speed ADC, a designer would also be required to design a second much more accurate ADC. Although the secondary ADC is designed to be slower, the constraint of higher accuracy makes the secondary ADC design non-trivial. Furthermore, the secondary ADC adds additional power and area consumption.

Another limitation of the approach of Fig. 2.11 is that the calibration time is a function of how slow the secondary ADC is. Thus while the secondary ADC can be made slower than the main ADC, the minimum speed (and thus minimum power required to implement the secondary ADC) is set by the desired calibration time, which as discussed in Sect. 2.3 should be as short as possible.

2.4.2 Split-ADC Gain Error Calibration

One topology which has proven to be highly effective in reducing calibration times in background schemes is the 'dual-ADC' or 'split-ADC' approach [10, 14, 18]. Shown in Fig. 2.12, the split-ADC takes a single ADC and splits it into two almost identical ADCs where each ADC has half the area, and half the thermal noise floor (thus half the power) of the overall ADC. The final ADC output is derived by taking the average of each ADC output – hence power and area of the split-ADC topology to a first order are not increased over a conventional ADC [18].

Each ADC is identical, except the residue transfer curve of the stage under calibration in one ADC is designed differently than the other. As a result when the

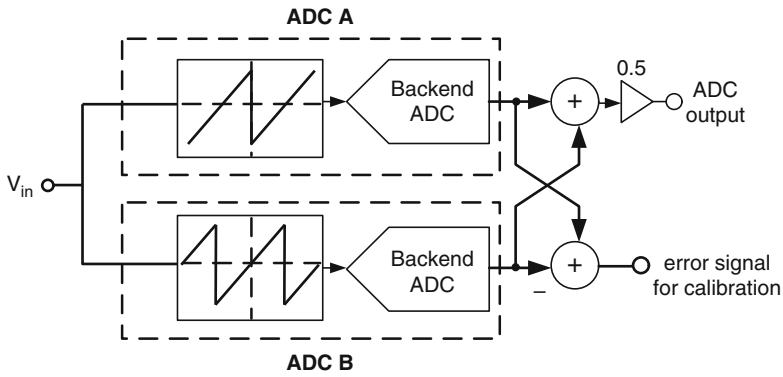


Fig. 2.12 Split-ADC topology

ADCs are free of errors both ADCs produce the same output. However when errors are included each ADC produces a different output. Since the analog input effectively appears as common mode to the split-ADCs, the error signal which is formed by the difference of the two ADCs is very weakly correlated with the analog input. However the error sources are very highly correlated with the difference in ADC outputs (i.e. error signal) due to each residue transfer curve being designed slightly differently between split ADCs [10]. Thus error sources can be estimated very quickly in the background by only looking at a small number of clock cycles of the error signal.

Examples of chips with measured results using the split-ADC approach for gain error calibration include: [14] where a dual ADC approach was used to realize a 0.18 um CMOS 5 MS/s ADC with 77 dB SFDR and 12 mW power consumption, where only 4,096 clock cycles were required to achieve calibration. In [18] a 16-b 1 MS/s ADC using a split-ADC approach was implemented in 0.25 um CMOS, where the power consumption was 105 mW, and calibration was achieved in only 10^4 clock cycles.

2.4.3 Rapid DAC and Gain Error Correction

The split-ADC technique described in Sect. 2.4.2, while very effective at reducing calibration time when only pipeline stage gain errors are corrected, does not address how to rapidly measure and correct DAC errors. In [20] a technique is presented which allows for the rapid measurement and correction of both gain and DAC errors in the first stage of a multi-bit pipelined ADC. Like the split-ADC approach, in [20] two ADCs (ADC A and ADC B) simultaneously process in parallel the same analog input as shown in Fig. 2.13.

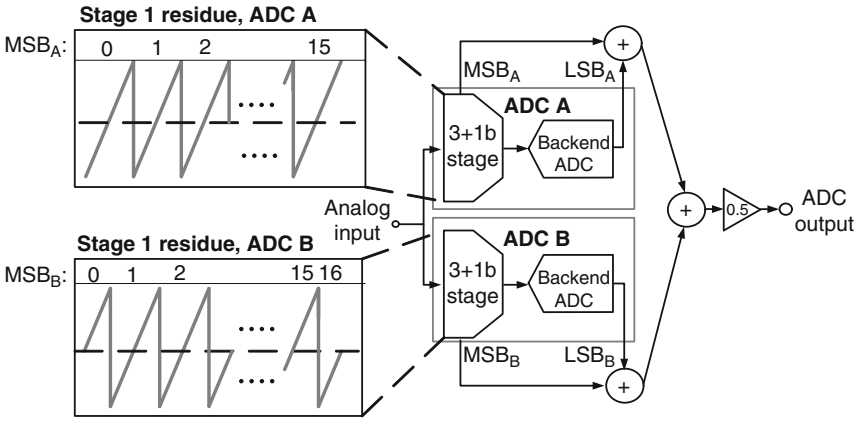


Fig. 2.13 Dual ADC topology of [20]

The final ADC output is generated by the average of the two ADC outputs, thus each ADC is designed with half the total capacitance, hence half the power and area of the overall ADC to meet thermal noise requirements. From Fig. 2.13 ADC A and B are identical except in each ADC the residue transfer function in the first stage is horizontally offset from the other by $^1/_2$ MSB.

If the analog input to the ADC is such that $MSB_A = i$, then MSB_B is either i or $i + 1$. The offset between the digital outputs of ADCs A and B for the range of analog inputs where $MSB_A = i$ and $MSB_B = i$ is denoted Δ_{i1}, and Δ_{i2} where $MSB_A = i$ and $MSB_B = i + 1$ as shown in Fig. 2.14.

In an ideal ADC $\Delta_{i1} = \Delta_{i2}$ (Fig. 2.14), however with DAC and/or gain errors the difference between Δ_{i1} and Δ_{i2} is precisely the error due to missing codes that occurs when MSB_B changes from i to $i + 1$ as shown in Fig. 2.15.

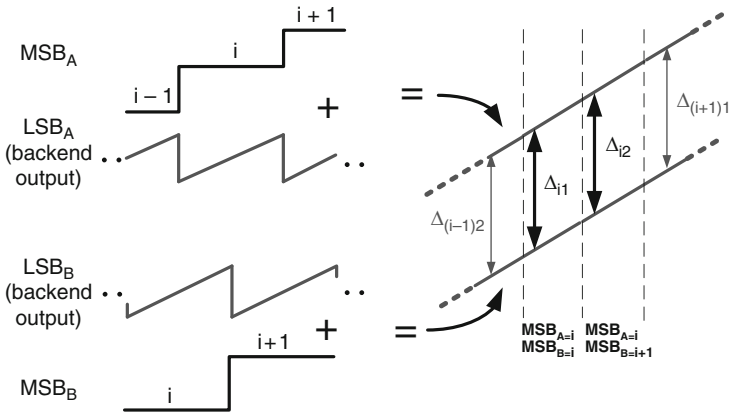


Fig. 2.14 Transfer curves of first stage (MSB), backend ADC (LSB) and total ADC outputs from each split ADC with no errors

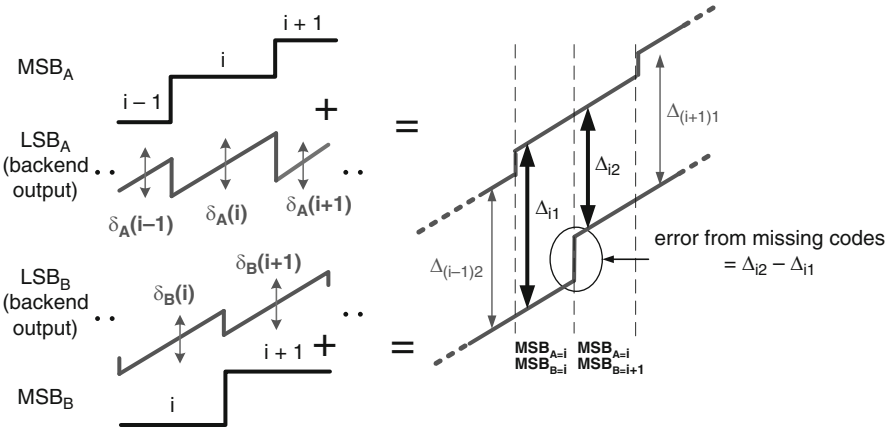


Fig. 2.15 Transfer curves of key ADC outputs with gain, DAC errors included

An accurate measure of Δ_{i1} and Δ_{i2} (thus accurate measure of error) can thus be made by simply measuring the average values of Δ_{i1}, Δ_{i2}, using a first order IIR filter with transfer function $\mu/[1-(1-\mu)z^{-1}]$.

In other words, the output of ADC A is used as an ideal reference for ADC B when $MSB_A = i$ to measure $\bar{\Delta}_{iB} = \bar{\Delta}_{i1} - \bar{\Delta}_{i2}$. In a similar manner the error due to missing codes at all other MSB transitions can be measured for ADC B. Errors due to missing codes for ADC A are measured by noting that $\Delta_{i2} - \Delta_{(i+1)1}$ is the error due to missing codes in ADC A when MSB_A changes from i to i + 1 as shown in Fig. 2.15. Hence the missing code errors in ADC A can be determined using already measured values $\bar{\Delta}_{i2} - \bar{\Delta}_{(i+1)1}$. Errors due to missing codes at all other MSB transitions in ADC A are measured using an identical extension as done for ADC B.

With the errors from missing codes at each MSB transition measured, each ADC is corrected by shifting each ADC's digital output as a function of MSB such that the overall transfer function of each ADC is free from missing codes due to errors in the first stage as shown in Fig. 2.16 (same done for ADC A).

Rapid background calibration is achieved as every analog input while $MSB_A = i$ produces outputs in ADCs A and B which when subtracted immediately give estimates of Δ_{i1} or Δ_{i2}. In contrast statistical techniques use statistical correlations which require many output samples to extract similar information. As long as the input is sufficiently busy to generate a sufficient number of estimates of Δ_{i1}, Δ_{i2}, for all i, there is no constraint on the type of input signal to the ADC.

It is noted that the approach of [20] is very similar to the background calibration technique of Sect. 2.4.1, where a slow but more accurate ADC is used in parallel to the ADC under calibration [19]. In the approach of [20] however, since the residue transfer function of one of the split ADCs is offset, ADC A does not suffer an error in the first stage for the same input as ADC B, thus one ADC can be used as an ideal reference for the other, eliminating the need for one of the ADCs to be more accurate than the other. Hence there is no need to trade higher accuracy with lower sampling rates in the second ADC; both ADCs can operate at the same

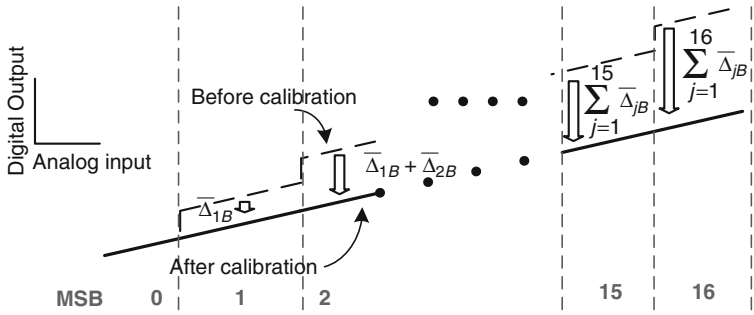


Fig. 2.16 Illustration of how correction terms for ADC B are derived from estimates of missing codes (correction topology of ADC A is identical)

speed, and both ADCs be used to digitize the analog input. Thus the power of the additional ADC also goes towards lowering the noise floor in the digital output, unlike [19] where the additional ADC (since it operates slower) only aids the correction scheme. Furthermore using the technique outlined in [20], both ADCs are calibrated whereas in [19] only one ADC is.

In [20], a chip was fabricated in 0.18 um CMOS, where at 45 MS/s the ADC was able to improve its SNDR/SFDR from 46.9/48.9 dB to 60.1/70 dB within only 10^4 clock cycles. For comparison, a statistical technique would require on the order of $2^{22} = \sim 4$ million clock cycles. Thus the approach of [20] shows a reduction of calibration time by more than two orders of magnitude. Figure 2.17 shows the INL before and after calibration, and Fig. 2.18 the improvement of ADC SNDR, SFDR with calibration cycles.

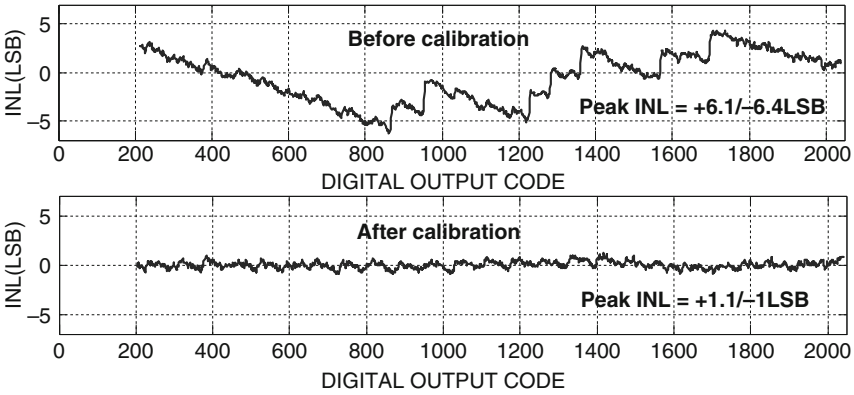


Fig. 2.17 INL before and after calibration

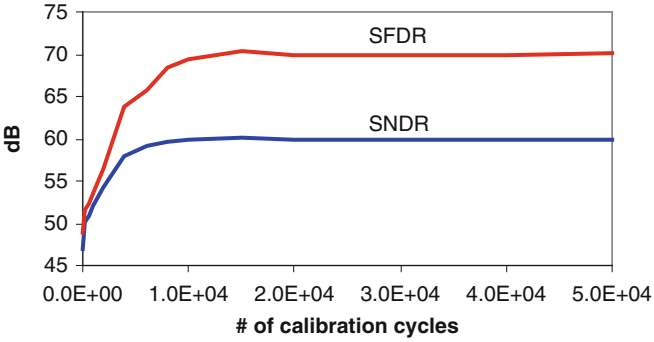


Fig. 2.18 SNDR, SFDR improvement with calibration cycles in [20]

2.5 Using Digital Calibration to Build Low Power 'Smart-ADCs'

In addition to enabling higher linearity, digital calibration enables relaxed analog circuitry and thus lowers power consumption. The vast majority of prior publications have exploited the relaxed analog requirements by using opamps with lower DC gain and thus lower power. Even with low DC gain however, opamps still consume large amounts of power in pipelined ADC. Recently approaches have emerged which heavily leverage the analog-digital tradeoff afforded by digital calibration by replacing opamps in closed-loop with topologies which are much less accurate in the analog domain, but significantly more power efficient. Sections 2.5.1 and 2.5.2 discuss a few examples of low-power 'smart-ADC' topologies.

2.5.1 Open Loop, Non-linear Gain Error Calibration

In [21] an open-loop technique is used in a 4-b first pipeline stage in a 12-b 75 MS/s ADC. Figure 2.19 illustrates the architecture of the 4-b pipeline stage used in [21].

Rather than using an opamp in closed-loop, an open-loop differential amplifier is used to achieve the desired gain of 8× in the pipeline stage, where the gain is set approximately by $g_m R_{load} = 8$, where g_m is the transconductance of the differential pair M1–M2, and R_{load} the load impedance seen by the differential pair.

The unity gain frequency of the amplifier in [21] is given by:

$$\omega_t = g_m / C_L \tag{2.1}$$

where ω_t is the unity gain frequency of the differential pair M1 – M2, and C_L the load capacitance seen at the output of the amplifier (i.e. nodes V_{outp}, V_{outn}). In contrast a closed-loop topology with a gain of 8× would have had at best a unity gain

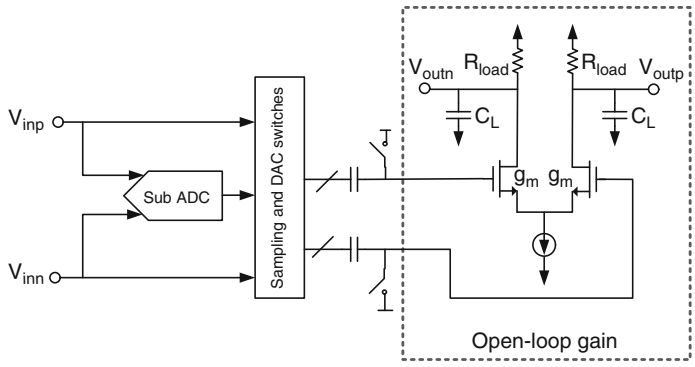


Fig. 2.19 4-b pipeline stage using open-loop amplifier

frequency 1/8th of that in Eq. 2.1. As a result it was shown in [21] that an open-loop approach yielded a power savings of 60% compared to a closed-loop opamp based approach.

In [21] low power was achieved by significantly relaxing the analog requirements and measuring and correcting the gain errors in the digital domain. In particular, in addition to the gain being set by $g_m R_{load}$ which can vary significantly due to temperature, the gain varies significantly as a function of the input to the pipeline stage as shown in Fig. 2.20.

To compensate for the nonlinearity and gain variation in [21], a calibration scheme was developed which measures and corrects the nonlinearity of the open-loop amplifier in the digital domain by using an inverse nonlinearity function, $f^{-1}(x)$, as shown in Fig. 2.21.

In [21] a statistics based digital background calibration scheme was used to estimate the value of $f^{-1}(x)$. It is noted that a spit-ADC technique to rapidly measure non-linear gain errors was proposed in [22].

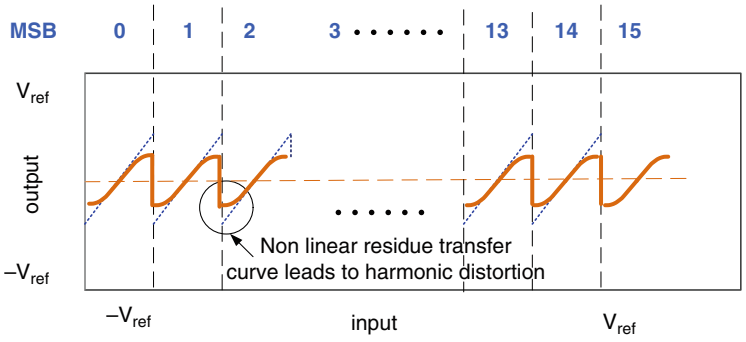


Fig. 2.20 Non-linear transfer curve for residue transfer function – ideal transfer curve shown in dashed lines

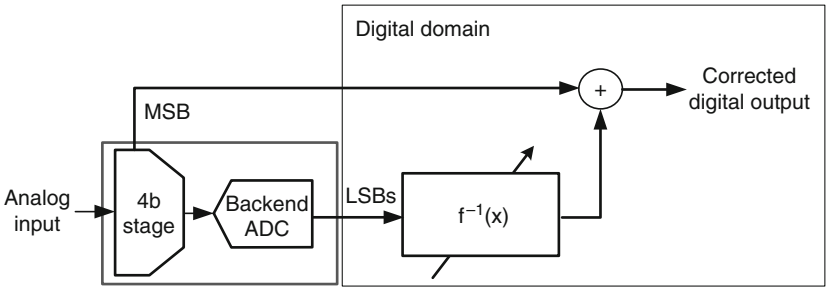


Fig. 2.21 Digital nonlinearity correction scheme used in [21]

2.5.2 Capacitive Charge Pump Based Pipelined ADC

In [23] a technique to achieve gain in pipelined stages using capacitive charge pumps is presented. The approach leverages digital calibration techniques to achieve a good linearity, yet achieve very low power (9.9 mW for 58 dB SNDR, at 50 MS/s).

In capacitive charge-pumps, successively larger voltages are attained by sampling a voltage on many capacitors in one clock phase, and connecting each capacitor in series in the next clock phase, i.e. gain is achieved by addition rather than multiplication [24]. In [24] an opamp based capacitive charge pump approach was used to achieve a low power algorithmic ADC. In [23] a capacitive charge pump approach is implemented in a pipelined ADC without opamps, and thus achieves even further power reduction. Figure 2.22 illustrates an example of how a gain of $\sim 2\times$ can be achieved using the approach discussed in [23].

A unity gain buffer is included to prevent charge sharing between sampling and load capacitors. Using the approach of Fig. 2.22, the classic gain-bandwidth trade-off which binds opamp based approaches is decoupled – gain is achieved by the capacitor arrangement, whereas the bandwidth of the output, V_{out}, is determined by the unity gain buffer and C_{load}. Thus in [23], a gain of $2\times$ can be achieved without compromising $2\times$ the bandwidth – as would otherwise be required in an opamp-based approach. Since bandwidth is approximately linearly related to power, the work of [23] enables a power reduction of at least 50% over opamp-based topologies.

To avoid amplifying offsets, the sampling network in [23] is arranged such that the differential input was sampled across the sampling capacitors, as shown in Fig. 2.23, which illustrates the 1.5-b pipeline stage topology used in [23].

It is noted that switch S0 was included in Fig. 2.23 to ensure bottom plate sampling by switches S1 and S2 (to minimize charge injection hence enable a high linearity). Since common-mode rejection is implemented in the sampling network

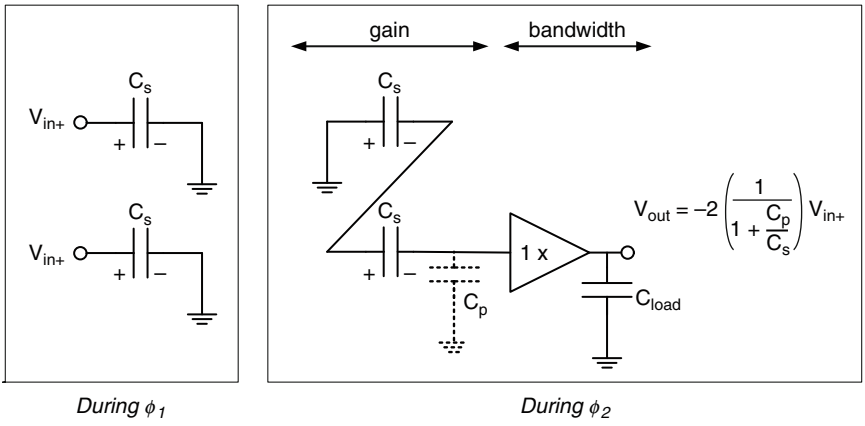


Fig. 2.22 Gain of near $2\times$ using a capacitive charge-pump approach

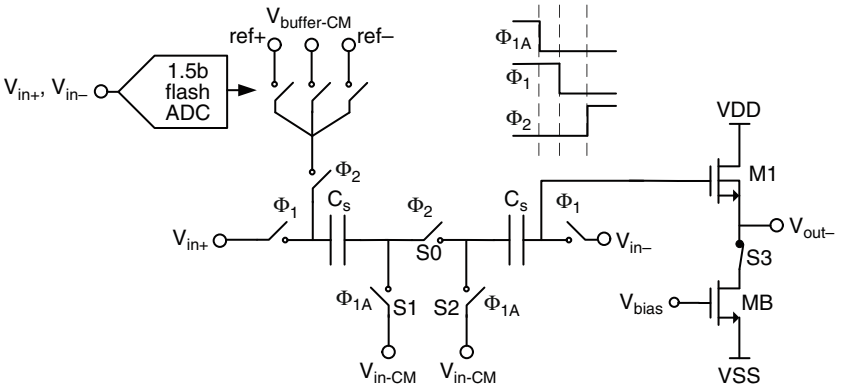


Fig. 2.23 Topology of each 1.5-b pipeline stage in [23] – the positive half is shown; the negative half is identical with a reversal of positive/negative signs

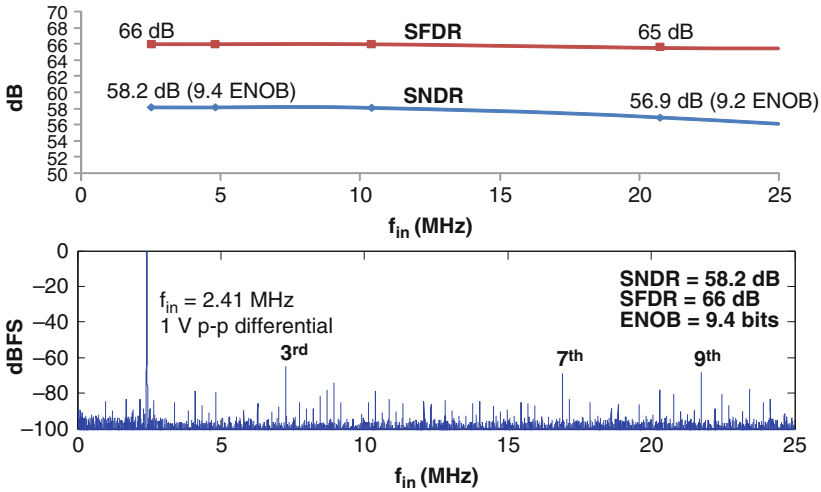


Fig. 2.24 SNDR and SFDR variation of [23] with input frequency, and FFT with $f_{in} = 2.4\,\text{MHz}$ (with $f_s = 50\,\text{MS/s}$)

rather than the buffers themselves, fully differential buffers are not required, thus the topology suppresses common-mode noise, yet requires no common-mode feedback.

The impact of parasitic capacitors were minimized in [23] by using the smallest switches possible to achieve the desired linearity and settling time. Sampling capacitors C_s were implemented with metal capacitors so that the gain was determined primarily by linear components. To account for the gain of each stage being smaller than the ideal 2×, standard digital gain error calibration as described in Sect. 2.3.2 was used to estimate and correct the large gain error of each pipeline stage. In [23]

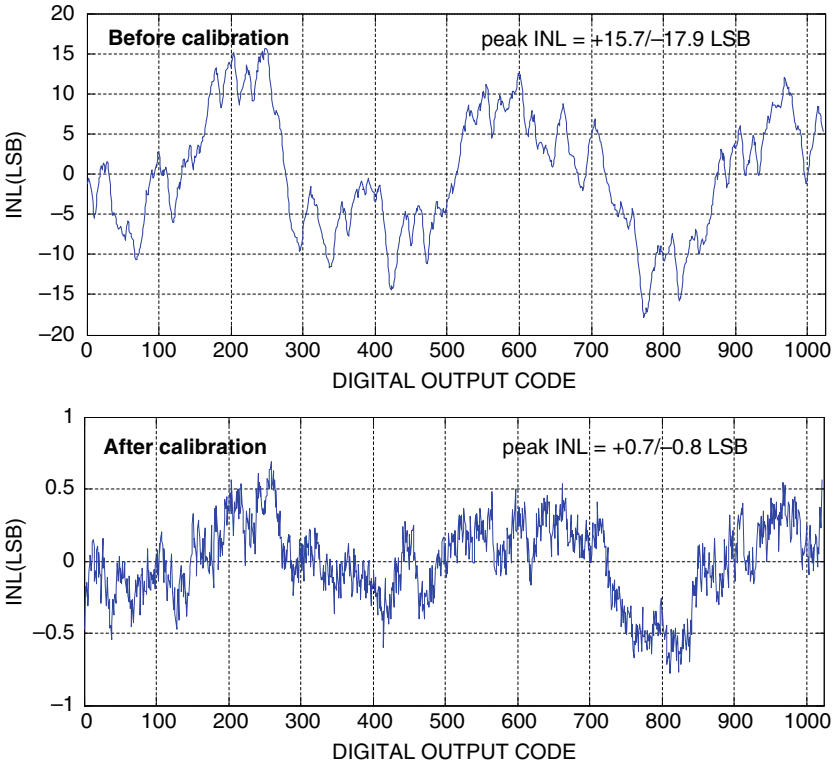


Fig. 2.25 INL before and after calibration (LSB at the 10-b level)

a foreground calibration approach was used to measure and correct the gain error of each pipeline stage, however background calibration could also have been used with the low power topology.

In [23] a prototype was implemented in a 1.8 V, 0.18 μm CMOS process. The total power of the ADC was 9.9 mW, including 3.9 mW from all active circuitry, and 6 mW from all clocking and clock distribution circuits. Measured results from the prototype of [23] showed a peak SNDR/SFDR of 58.2/66 dB, and a peak ENOB of 9.4-b as shown in Fig. 2.21.

Figure 2.6 shows the INL of the ADC before and after calibration – clearly illustrating the significant benefit of digital calibration in enabling a significant improvement in linearity, while enabling very low power consumption in the analog domain.

Using a figure-of-merit of Power/($2^{ENOB} \times f_s$), the ADC of [23] achieved 0.3 pJ/step. As shown in Fig. 2.26, the topology of [23] is amongst the most power efficient 10-b publications in the 5–80 MS/s space, thus displaying the beneficial impact digital calibration can have on enabling low power 'smart-ADCs'.

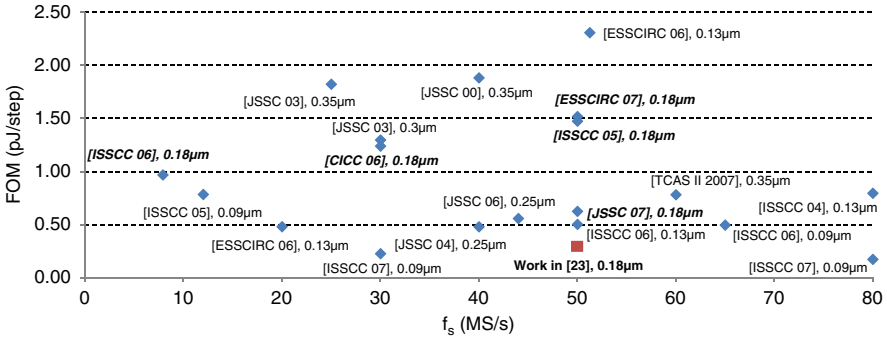


Fig. 2.26 FOM comparison with other 10-b ADCs between 5–80 MS/s (0.18 µm papers in bold italics)

2.6 Summary

In this paper a review of common pipelined ADC calibration techniques was given, with a discussion of the associated tradeoffs of each approach. Foreground calibration was shown to have a simple topology, however requires the ADC to be taken offline. Statistics based background calibration was shown to be continuously functional, however at the penalty of lengthy calibration times, which are undesirable in an industrial environment. Split-ADC based techniques were shown to be the most effective approaches published thus far to enable substantial calibration time reduction. A brief discussion of 'smart-ADCs' which leverage digital calibration techniques for substantial reductions in analog complexity and power were also reviewed, where open-loop amplifiers and charge pump based gain techniques were shown to be promising strategies to achieve low power consumption.

References

1. K. Bult, G.J.G.M. Geelen, "A fast settling CMOS op amp for SC circuits with 90-dB DB Gain", *IEEE Journal of Solid-State Circuits*, vol.25, pp.1379–1384, Dec. 1990.
2. F. You, S.H.K. Embabi, E. Sanchez-Sinencio, "Multistage amplifier topologies with nested G_m-C compensation," *IEEE Journal of Solid-State Circuits*, vol.32, no.12, pp.2000–2011, Dec. 1997.
3. S.-H. Lee, B.-S. Song, "Simplified digital calibration for multi-stage analog-to-digital converters," *1993 IEEE International Symposium on Circuits and Systems, 1993, ISCAS '93*, vol.2, 3–6, pp.1216–1219, May 1993.
4. D.Y. Chang, J. Li, U.K. Moon, "Radix-based digitial calibration techniques for multi-stage recycling pipelined ADCs", *IEEE Transactions on Circuits and Systems I*, vol.51, pp.2133–2140, Nov. 2004.
5. C.R. Grace, P.J. Hurst, S.H. Lewis, "A 12b 80 MS/s pipelined ADC with bootstrapped digital calibration", in *IEEE International Solid-State Circuits Conference. (ISSCC) Digital Technical Papers*, pp.460–539, Feb. 2004.

6. U.K. Moon, B.-S. Song, "Background digital calibration techniques for pipelined ADC's", *IEEE Transactions on Circuits and Systems II*, vol.44, pp.102–109, Feb. 1997.
7. S.-U. Kwak, B.-S. Song, K. Bacrania, "A 15-b, 5-Msample/s low-spurious CMOS ADC," *IEEE Journal of Solid-State Circuits*, vol.32, no.12, pp.1866–1875, Dec. 1997.
8. O.E. Erdogan, P.J. Hurst, S.H. Lewis, "A 12-b digital-background-calibrated algorithmic ADC with -90-dB THD," *IEEE Journal of Solid-State Circuits*, vol.34, no.12, pp.1812–1820, Dec. 1999.
9. I. Galton, "Digital cancellation of D/A converter noise in pipelined A/D converters," *IEEE TCAS-II*. vol.47, pp.185–196, Mar. 2000.
10. J. Li, U.K. Moon, "Background calibration techniques for multistage pipelined ADCs with digital redundancy," *IEEE Transactions on Circuits and Systems II*, vol.50, pp.531–538, Sep. 2003.
11. Y. Chiu, C.W. Tsang, B. Nikolic, P.R. Gray, "Least mean square adaptive digital background calibration of pipelined analog-to-digital converters", *IEEE TCAS-I*, vol.51, pp.38–46, Jan. 2004.
12. X. Wang, P.J. Hurst, S.H. Lewis, "A 12-bit 20-Msample/s pipelined analog-to-digital converter with nested digital background calibration," *IEEE Journal of Solid-State Circuits*, vol.39, no.11, pp.1799–1808, Nov. 2004.
13. E. Siragusa, I. Galton, "A digitally enhanced 1.8-V 15-bit 40-MSample/s CMOS pipelined ADC," *IEEE Journal of Solid-State Circuits*, vol.39, no.12, pp. 2126–2138, Dec. 2004.
14. J. Li, G.-C. Ahn, D.-Y. Chang, U.-K. Moon, "A 0.9-V 12-mW 5-MSPS algorithmic ADC with 77-dB SFDR," *IEEE Journal of Solid-State Circuits*, vol.40, no.4, pp.960–969, Apr. 2005.
15. H.-C. Liu, Z.-M. Lee, J.-T. Wu, "A 15-b 40-MS/s CMOS pipelined analog-to-digital converter with digital background calibration," *IEEE Journal of Solid-State Circuits*, vol.40, no.5, pp.1047–1056, May 2005.
16. S. Ray, B.-S. Song, "A 13-b linear, 40-MS/s pipelined ADC with self-configured capacitor matching", *IEEE JSSC*, vol.42, pp.463–474, Mar. 2007.
17. E. Siragusa, I. Galton, "A digitally enhanced 1.8-V 15-bit 40-MSample/s CMOS pipelined ADC," *IEEE Journal of Solid-State Circuits*, vol.39, no.12, pp.2126–2138, Dec. 2004.
18. J. McNeill, M. Coln, B. Larivee, "A split-ADC architecture for deterministic digital background calibration of a 16b 1 MS/s ADC," *Solid-State Circuits Conference, 2005. Digest of Technical Papers. ISSCC. 2005 IEEE International*, vol.48, 6–10, pp.276–598, Feb. 2005.
19. Y. Chiu, C.W. Tsang, B. Nikolic, P.R. Gray, "Least mean square adaptive digital background calibration of pipelined analog-to-digital converters," *IEEE TCAS-I*, vol.51, pp.38–46, Jan. 2004.
20. I. Ahmed, D.A. Johns, "An 11-Bit 45 MS/s pipelined ADC with rapid calibration of DAC errors in a multibit pipeline stage," *IEEE Journal of Solid-State Circuits*, vol.43, no.7, pp.1626–1637, Jul. 2008.
21. B. Murmann et al., "A 12-bit 75-MS/s pipelined ADC using open-loop residue amplification," *IEEE Journal of Solid-State Circuits*, vol.38, pp.2040–2050, Dec. 2003.
22. J.A. McNeill, S. Goluguri, A. Nair, "Split-ADC" digital background correction of open-loop residue amplifier nonlinearity errors in a 14b pipeline ADC," *IEEE International Symposium on Circuits and Systems, 2007. ISCAS 2007*, vol., no. 27–30, pp.1237–1240, May 2007.
23. I. Ahmed, J. Mulder, D.A. Johns, "A 50MS/s 9.9 mW pipelined ADC with 58dB SNDR in 0.18um CMOS using capacitive charge-pumps," *Solid-State Circuits Conference, 2009. Digest of Technical Papers. ISSCC. 2009 IEEE International*, vol.52, 9–12, pp.164–165, Feb. 2009.
24. P. Quinn, M. Pribytko, "Capacitor matching insensitive 12-b 3.3 MS/s algorithmic ADC in 0.25/spl mu/m CMOS," *Custom Integrated Circuits Conference, 2003. Proceedings of the IEEE 2003*, vol.21–24, pp.425–428, Sep. 2003.

Chapter 3
High-Resolution and Wide-Bandwidth CMOS Pipeline AD Converters

Hans Van de Vel

Abstract High-resolution wide-bandwidth ADCs in nm-CMOS are key enablers in increasing the level of digitization and integration in cellular base station receivers. This paper discusses smart techniques to overcome the limitations of low supply voltage and low intrinsic device gain. A 14 b 100 MS/s ADC in 90 nm CMOS is described demonstrating that good power efficiency can be achieved in nm-CMOS with a low supply voltage.

3.1 Introduction

Reducing cost in wireless infrastructure systems is the main driver in pushing CMOS analog-to-digital converters towards higher resolutions and wider bandwidths. First the cost of a cellular base station is reduced by adopting a higher level of digitization in the receiver. Such a highly digitized, multi-channel receiver pushes the ADC towards higher dynamic range and higher speed. Second the cost can be further reduced by adopting a higher level of integration, requiring the ADC to be implemented in nanometer-CMOS technology.

Figure 3.1 shows the block diagram of a typical multi-channel receiver, where the extraction of individual channels is implemented in the digital domain. The antenna signal is first filtered, amplified and down-converted, and subsequently the complete cellular frequency band with a multiple of channels is digitized by the ADC. The digital channelizer then performs a set of filtering and down-conversion operations to extract the individual channels for further processing. The bandwidth of the cellular band at the input of the ADC is in the order of tens of MHz. The system's dynamic range requirements dictate the ADC's noise level and linearity. For 2.5 G and 3 G cellular standards like EDGE and UMTS, an ADC with a signal-to-noise ratio (SNR) of 72–75 dB and a spurious-free dynamic range (SFDR) of 85–90 dB is required. For a GSM system, the difference in channel attenuation for near and far

H. Van de Vel (✉)
NXP Semiconductors, Eindhoven, The Netherlands
e-mail: hans.van.de.vel@nxp.com

A.H.M. van Roermund et al. (eds.), *Analog Circuit Design: Smart Data Converters, Filters on Chip, Multimode Transmitters*, DOI 10.1007/978-90-481-3083-2_3,
© Springer Science+Business Media B.V. 2010

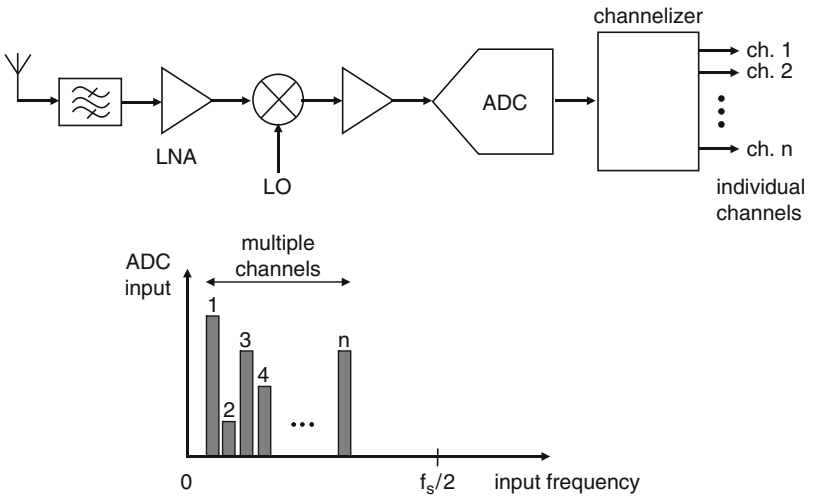


Fig. 3.1 Typical multi-channel receiver

users demands an even higher SFDR of 100 dB. The ADC's resolution should then be 14 b or higher such that the errors and spurs arising due to the quantization are at least 10 dB lower than the required maximum level.

The blocks in Fig. 3.1 can be partitioned over a number of dies or integrated on a single die. Integration on a single die brings a significant reduction of the total system cost. Since a large part of the signal processing is performed in the digital domain, it is advantageous in terms of power and area to implement the receiver in the most advanced nanometer-CMOS technology. As a key enabler for this single-die approach, the realization of high-resolution and wide-bandwidth ADCs in nm-CMOS is the focus of this paper.

Recently AD converters with resolutions of 14 b or higher, and bandwidths of 50 MHz or higher, have achieved good power efficiency in 0.18 μm-CMOS technology with a 3 V supply voltage [1]. Such converters use high-gain amplifiers to achieve high linearity, and large signal swings for good noise performance. In nm-CMOS, both linearity and noise performance are compromised by respectively low intrinsic transistor voltage gains and low signal swings due to the low supply voltage. This calls for a smart approach [2] to cope with the challenges of nm-CMOS design and to benefit from the excellent digital capabilities. This paper discusses digital calibration of non-linearity, range-scaling and an SHA-less architecture, being key enabling techniques for such an approach. It is then demonstrated how these techniques can be used for a power-efficient realization of a 14 b 100 MS/s ADC in a 1.2 V 90 nm CMOS technology [3, 4].

The dominant architecture for high-resolution wide-bandwidth CMOS ADCs is the pipeline converter [1, 3–7]. This paper assumes an opamp-based switched-capacitor implementation. Promising alternative pipeline topologies include open-loop amplifier-based [8], comparator- and zero-crossing-based [9] and charge

domain [10] implementations. Typically very good power efficiency is achieved, but additional effort is required to increase the robustness [11]. Furthermore not every alternative pipeline topology easily delivers the combination of high resolution and wide bandwidth.

3.2 Digital Calibration of Non-Linearity

Digital calibration is an attractive solution to overcome linearity limitations, while taking advantage of the high density and low power consumption of digital circuits in nm-CMOS. The additional design complexity and the extra power consumption of digital calibration circuits should be justified by significant power savings in the analog circuits of the ADC.

In an ADC non-linearity may result from various sources of errors. These errors can be either static or dynamic. This paper discusses digital calibration of static non-linearity only.

Figure 3.2 shows a general block diagram of a pipeline ADC, which basically is a chain of pipeline stages, where each stage resolves a certain number of bits and generates an amplified residue that is digitized by the next stages. Each pipeline stage consists of an analog-to-digital sub-converter (ADSC), a sample-and-hold circuit

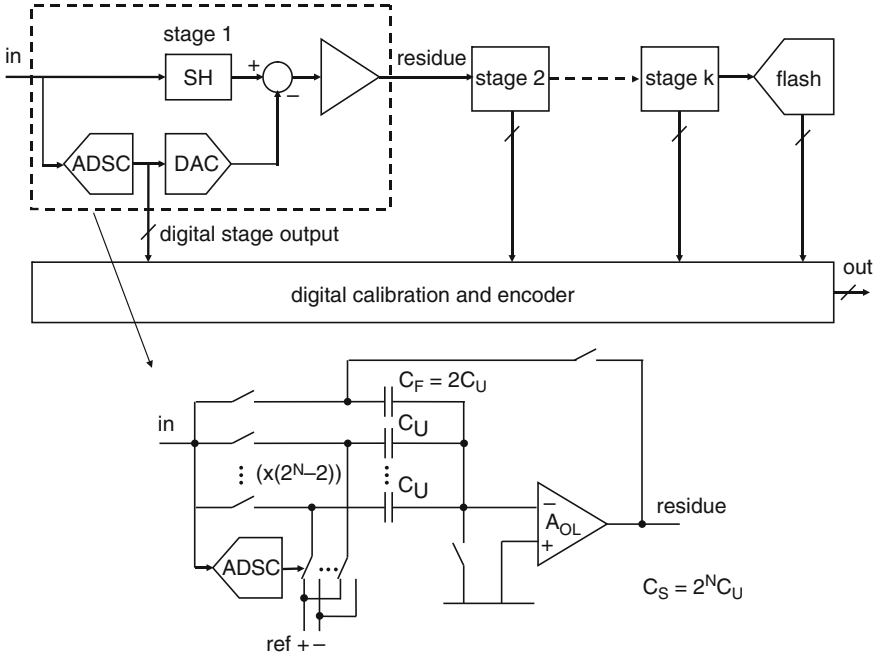


Fig. 3.2 General pipeline ADC architecture

(SH), a digital-to-analog converter (DAC), a subtractor and an amplifier. In CMOS technology the latter four functions usually are combined in a switched-capacitor (SC), multiplying DAC-based (MDAC) gain stage [12].

The dominant limitations in the static linearity of an SC pipeline stage are comparator offset in the ADSC, DAC non-linearity and stage gain non-linearity. The effect of comparator offset in the ADSC is mitigated in the digital domain by using over-range codes [5]. The gain of a typical stage resolving N bits is then 2^{N-1}, and half of the residue signal range is over-range. Sacrificing part of the conversion range is justified by the power saving in the ADSC. Following subsections discuss the DAC and stage gain non-linearity.

3.2.1 DAC Non-linearity

The static DAC non-linearity in an SC pipeline stage is mainly caused by capacitor mismatch. In [6] it is shown that the differential non-linearity (DNL) due to the capacitor mismatch of the first stage is inversely proportional to the square root of the total capacitance value in the DAC. For the pipeline stage in Fig. 3.2 the DNL, normalized to the least significant bit (LSB), is:

$$DNL = \frac{\Delta C_U}{C_U} 2^{N_T - N + 1} \tag{3.1}$$

where C_U is the MDAC's unit capacitance and N_T and N are respectively the total resolution of the ADC and the resolution of the first stage. The standard deviation of the relative capacitor mismatch is:

$$\sigma_{\frac{\Delta C_U}{C_U}} = \frac{A_C}{\sqrt{C_U}} \tag{3.2}$$

where A_C is a technology-dependent proportionality constant. Taking three times the standard deviation as a worst-case value for the capacitor mismatch, the DNL is:

$$DNL = \frac{3A_C 2^{N_T - N/2 + 1}}{\sqrt{C_S}} \tag{3.3}$$

where $C_S = (2^N - 2)\, C_U + C_F = 2^N C_U$ is the total sampling capacitance value of the MDAC and $C_F = 2C_U$ is the stage's feedback capacitance. As a numerical example consider $A_C = 0.25\%\ [\sqrt{fF}]$, $N_T = 14$ and $N = 4$. A 4 pF sampling capacitance is then needed in the first stage, to achieve a DNL smaller than an LSB.

The DAC non-linearity can thus be minimized by choosing a large value for the first stage sampling capacitance, which on the one hand is also beneficial for the noise performance, but on the other hand reduces speed and sampling linearity and increases the loading of the ADC driver. Digital calibration of DAC non-linearity

enables the use of smaller capacitance values. Since the variation of capacitor mismatch with temperature or aging in most cases is insignificant, the calibration can be performed with a foreground technique, e.g. [13], in a time-period during test or at start-up. If this time-period is not available, then the calibration has to rely on a background technique like the one proposed in [14].

3.2.2 Stage Gain Non-linearity

The dominant contributions to the stage gain non-linearity are a first-order gain error and a third-order gain compression. The closed-loop gain (A_{CL}) of a typical SC pipeline stage (Fig. 3.2) that resolves N bits is:

$$A_{CL} = \frac{C_S}{C_F} \cdot \frac{1}{1 + \frac{1}{kA_{OL}}} \cong \frac{C_S}{C_F} - \frac{C_S}{C_F} \cdot \frac{1}{kA_{OL}} \tag{3.4}$$

where A_{OL} is the open-loop gain of the stage's opamp, and feedback factor k is:

$$k = \frac{C_F}{C_S + C_P} \tag{3.5}$$

where C_P is the parasitic capacitance at the amplifier's input node.

If the opamp gain is high enough, the second term in Eq. 3.4 is negligible and the stage gain is determined by a ratio of capacitors. This second term times half the sub-range then needs to be smaller than half an LSB. For this A_{OL} needs to be larger than $2^{N_T-N+1}/k$. For example consider $N_T = 14$, $N = 4$ and $k = 0.1$, then $A_{OL} > 86\,\text{dB}$. Equation 3.4 then becomes $A_{CL} = C_S/C_F$. Given that $C_S = 2^N C_U$ and $C_F = 2C_U$, then the closed-loop gain is 2^{N-1} which is required for an N-bit stage with an over-range equal to half of the residue signal range. A first-order gain error can then arise due to capacitor mismatch only and can be corrected by a digital calibration algorithm. The same considerations as in Sect. 3.2.1 hold in choosing either a foreground [13] or a background technique [4].

Such a high opamp gain is difficult to achieve in nm-CMOS due to the low intrinsic transistor gain and the low supply voltage. For example the single-stage telescopic opamp with gain-boosting, as used in [1], fails to deliver the combination of high gain and high signal swing in nm-CMOS. With a low opamp gain the stage gain is determined by both terms in Eq. 3.4, and consequently depends on the open-loop gain A_{OL} and the feedback factor k. Typically the feedback capacitance is tuned in the design phase – cf. Sect. 3.3 – such that the nominal value of A_{CL} equals 2^{N-1}. However a first-order gain error can arise due to capacitor mismatch, an inaccurate estimation of parasitic capacitances, inaccurate process characterization data or variation with temperature or aging. The first three effects can be corrected by a digital foreground calibration algorithm, while the fourth effect requires a background routine.

Next to a first-order gain error, also a third-order gain compression can cause stage gain non-linearity. Assuming a single-stage opamp is used, the third-order gain compression is due to the input pair's transconductance and the output resistance. The third-order harmonic distortion (HD3) due to the input pair's transconductance is proportional to the opamp's output signal swing (V_{sig}) squared and inversely proportional to A_{OL} squared times V_{GT} squared [15]:

$$HD3 = \frac{V_{sig}^2}{128 A_{OL}^2 V_{GT}^2} \tag{3.6}$$

where $V_{GT} = V_{GS} - V_T$ is the overdrive voltage of the input pair's transistors. The HD3 due to the output resistance is also proportional to V_{sig} squared. Assuming Eq. 3.6 determines the stage's HD3, then the effect of the third-order gain compression is smaller than half an LSB, if A_{OL} is larger than $2^{(N_T-N-5)/2} \cdot V_{sig}/V_{GT}$. For example consider $N_T = 14$, $N = 4$, $V_{sig} = 1\,\text{V}$ and $V_{GT} = 0.2\,\text{V}$, then $A_{OL} > 29\,\text{dB}$. Thus, for very low opamp gains, also the third-order gain error needs to be digitally calibrated with a foreground or background [8] routine.

3.3 Range-Scaling in the First Pipeline Stage

The noise performance of nm-CMOS pipeline ADCs is compromised by signal swing limitations due to the low supply voltage. This section discusses range-scaling in the first pipeline stage as an effective technique to maximize the voltage efficiency over the pipeline chain. The range-scaling enables the realization of a power efficient noise-limited pipeline ADC.

3.3.1 Power Consumption in a Noise-Limited ADC

The power consumption of a noise-limited ADC is proportional to the SNR and the sampling rate:

$$P \propto \frac{kT \cdot SNR \cdot f_s}{\eta_{vol} \cdot \eta_{cur}} \tag{3.7}$$

where $\eta_{vol} = V_{sig}/V_{dd}$ is the voltage efficiency and η_{cur} is the current efficiency [16,17], k is Boltzmann's constant, T is the temperature and V_{dd} is the supply voltage. There is no increase in the power consumption in a noise-limited ADC with decreasing supply voltages, if the product of the voltage and the current efficiency $\eta_{vol}\eta_{cur}$ can be maintained constant.

The current efficiency η_{cur} can be maximized by using power-efficient single-stage opamps in the pipeline stages, as discussed above, and by implementing an SHA-less architecture, as discussed in Sect. 3.4.

The voltage efficiency η_{vol} on the other hand can be maximized by using a range-scaling first stage. The range-scaling is effective in reducing the power consumption of the stage's opamp [18], and decouples the choice of the stage's input and output signal swing, adding an extra degree of freedom.

3.3.2 Circuit Implementation

Figure 3.3 shows a circuit implementation of a range-scaling first pipeline stage, resolving four bits ($N = 4$). Only 1 out of 14 comparators is shown in the figure. The gain of a 4 b stage without range-scaling and with an over-range equal to half of the residue signal range is 8 ($= 2^{N-1}$). Then the input signal swing equals the output signal swing. In the range-scaling stage, a feedback capacitance C_{F2} is used to implement the range-scaling.

The operation of the stage is explained next. During phase φ_1 the charge in feedback capacitance C_{F2} is reset, and the voltage over sampling capacitances C_S and C_{SL1} to C_{SL14} tracks the input voltage, which is sampled at the falling edge of φ_{1e}. During phase φ_2 a capacitance sized $C_S/8$ and capacitance C_{F2} are connected in a feedback configuration over the opamp, and the sampling capacitances C_{SL1} to C_{SL14} are connected to the reference levels generated by a resistor ladder. At the falling edge of φ_{latch} the differences between the input signal and the reference

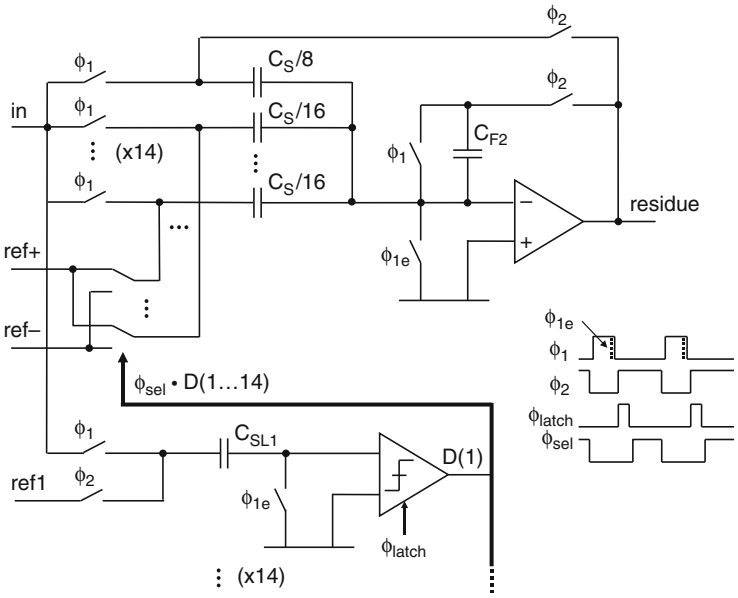


Fig. 3.3 Range-scaling first pipeline stage

levels are latched and the unit capacitances sized $C_S/16$ are either connected to plus or minus the reference voltage.

The stage's closed-loop gain then is:

$$A_{CL} = \frac{C_S}{C_S/8 + C_{F2}} \cdot \frac{1}{1 + \frac{1}{kA_{OL}}} \tag{3.8}$$

and the feedback factor k is:

$$k = \frac{C_S/8 + C_{F2}}{C_S + C_{F2} + C_P} \tag{3.9}$$

When C_{F2} equals zero, Eqs. 3.8 and 3.9 correspond to Eqs. 3.4 and 3.5.

The stage's gain is now also determined by the value of feedback capacitance C_{F2}. If the gain of the 4 b stage is smaller than 8, then the stage's input signal swing is larger than its output signal swing, which is typically limited due to the low supply voltage. For example, if the product kA_{OL} is very large and $C_{F2} = C_S/8$, then the gain of the 4 b stage equals 4. The input signal swing is then two times larger than the output signal swing, or correspondingly the range scales with a factor two.

3.4 SHA-Less Architecture

Often a dedicated input sample-and-hold amplifier (SHA) is used such that the input to the first pipeline stage is a sampled-data signal [6]. Such an SHA samples wide-bandwidth noise prior to any amplification and its power consumption is dictated by the noise, linearity and speed requirements. This section discusses an SHA-less architecture. Omitting the dedicated input SHA enables significant power savings, but poses additional constraints on the design of the first pipeline stage, as will be explained next.

If the first pipeline stage is not preceded by a dedicated input SHA, then the bandwidth and timing mismatch between the sampling operations in the MDAC and in the ADSC need to be minimized to minimize the aperture error. In the pipeline stage in Fig. 3.3, this aperture error is minimized by sampling simultaneously at the falling edge of φ_{1e} and by matching the track phase bandwidth. Since the input signal to the stage is a time-continuous signal, distortion at high frequency and high signal swing is reduced by bootstrapping the input switches in Fig. 3.3.

Because the range-scaling first pipeline stage (Fig. 3.3) is not preceded by a dedicated input SHA, slow settling in the track phase can cause non-linearity. At the onset of phase φ_1 the voltage over sampling capacitance C_S needs to settle to the instantaneous value of the input voltage. This settling behavior is characterized by a time constant $\tau_{track} = (R_S + R_{on,in} + R_{on,x}) \cdot C_S$, where R_S is the output resistance of the ADC driver and $R_{on,in}$ and $R_{on,x}$ are the on-resistances of the input

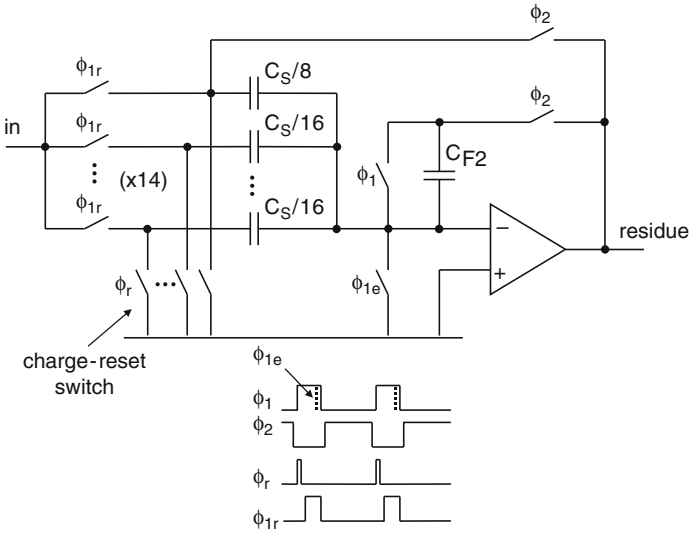


Fig. 3.4 Charge-reset switch in first stage

switch and the sampling switch respectively. The on-resistance of the switches can be as low as a few Ohm, but R_S typically is 50 Ω resulting in slow settling.

This slow settling causes inter-symbol interference (ISI), where the sampled voltage is related to the previously sampled voltage. For the circuit in Fig. 3.3 this relation is non-linear, because a non-linear part of the sampled charge in C_S is transferred to feedback capacitance C_{F2}, and is reset in phase φ_1. The ISI then results in distortion.

Implementing a charge-reset switch, as shown in Fig. 3.4, eliminates the ISI and the associated distortion mechanism. The charge-reset switch is on during reset phase φ_r, and the sampling capacitance C_S is connected to the input signal after the reset phase, in phase φ_{1r}. The settling behavior in the reset phase is characterized by $\tau_{track} = (R_{on,r} + R_{on,x}) \cdot C_S$, where $R_{on,r}$ is the on-resistance of the charge-reset switch. This settling is fast and the reset phase can be short.

3.5 A 1.2 V 14 b 100 MS/s ADC in 90 nm CMOS

The smart approach to designing high-resolution and wide-bandwidth nm-CMOS pipeline ADCs is demonstrated in this section. A 14 b 100 MS/s digitally calibrated pipeline ADC has been realized in a 90 nm CMOS technology with a 1.2 V supply voltage [3,4]. The ADC incorporates digital calibration of first- and third-order stage gain errors, a range-scaling first pipeline stage and an SHA-less architecture with a charge-reset switch.

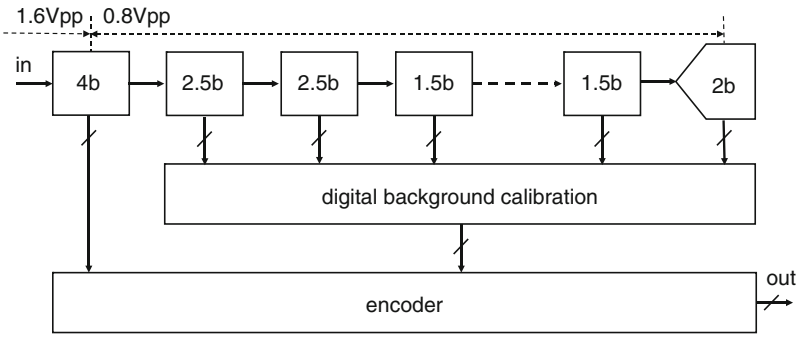


Fig. 3.5 14 b ADC architecture

3.5.1 ADC Architecture

Figure 3.5 shows the 14 b SHA-less ADC architecture. The 4 b first pipeline stage is followed by two 2.5 b stages, seven 1.5 b stages and a final 2 b flash stage. In the first pipeline stage, range-scaling decouples the choice of the stage's input and output signal swing, and both can be optimized separately. With a 1.2 V supply voltage, the peak-to-peak differential input signal swing can be as high as 1.6 V. The 0.2 V voltage headroom is sufficient for the on-chip reference buffers, since these buffers drive a static reference level. The first stage's opamp and the back-end pipeline chain on the other hand require a larger voltage headroom. The range-scaling first stage therefore reduces the output signal swing with a factor of 2 to 0.8 V. As discussed above, the first pipeline stage contains the charge-reset switch to eliminate the ISI-induced distortion mechanism.

The calibration block diagram is shown in Fig. 3.2. The first stage's analog residue voltage is digitized by the back-end pipeline chain and fed to the digital post-processing block. The digital correction block implements the inverse of the stage gain non-linearity in the first stage, using a third-order power series:

$$D_{res,corr} = b_1 \cdot D_{res} + b_3 \cdot D_{res}^3 \tag{3.10}$$

where the first- and third-order coefficients b_1 and b_3 are the correction parameters. The stage gain non-linearity is only corrected for the first stage, since for the later stages its effect is suppressed by the preceding gain, and the opamps have sufficient open-loop gain. It has been found empirically that the optimal value for the correction parameter b_1 is different from sample to sample, whereas for b_3 it is constant. Therefore only b1 is iteratively updated by the digital error estimation block, reducing the design complexity of the digital calibration algorithm. More details on the digital background error estimation can be found in [4] (Fig. 3.6).

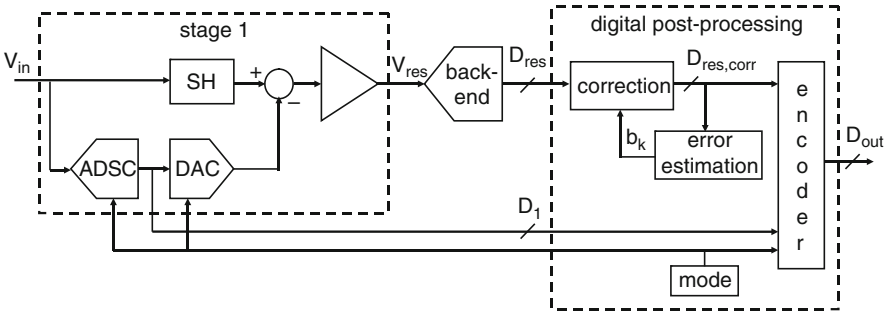


Fig. 3.6 Calibration block diagram

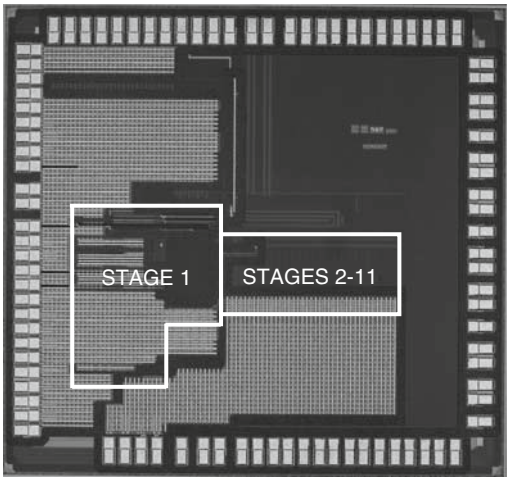


Fig. 3.7 Chip micrograph

3.5.2 Measured Results

The ADC has been fabricated in a baseline 90 nm CMOS process with one layer of poly and seven layers of metal (Fig. 3.7). It occupies an active chip area of 1 mm^2 and consumes 250 mW from a 1.2 V supply. This power figure includes the 50 mW power consumption in the on-chip reference buffers, and excludes the digital output buffers. All digital post-processing has been implemented off-chip on a PC. The simulated power consumption of an on-chip implementation of the digital calibration algorithm is 2 mW.

The differential non-linearity (DNL) and the INL measurements were done using a code-density test with a full-scale 4.3 MHz sinusoidal input signal. Figures 3.8 and 3.9 shows the DNL and INL before and after calibration. The missing codes in Fig. 3.8a and the large steps in Fig. 3.9a are the signatures of a large first-order gain error. Before calibration the INL is 133 LSB, which is ten times larger than

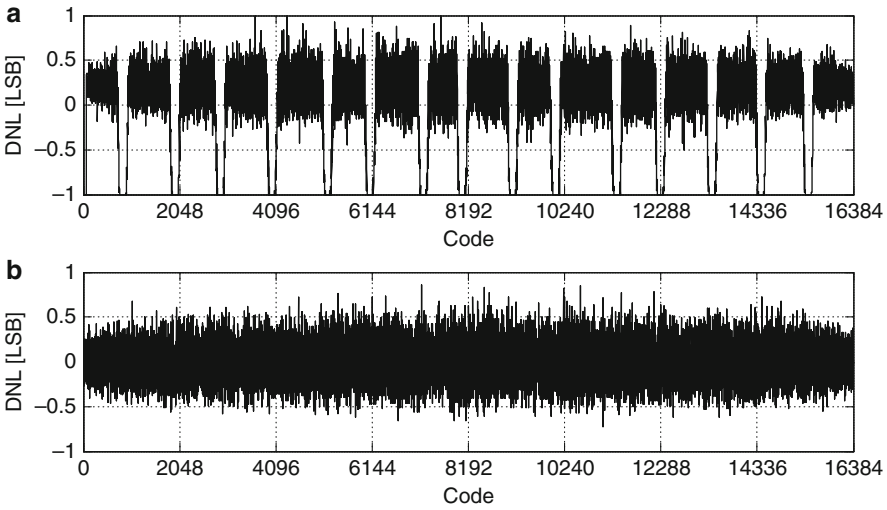


Fig. 3.8 8 Measured DNL. **(a)** Before calibration. **(b)** After calibration

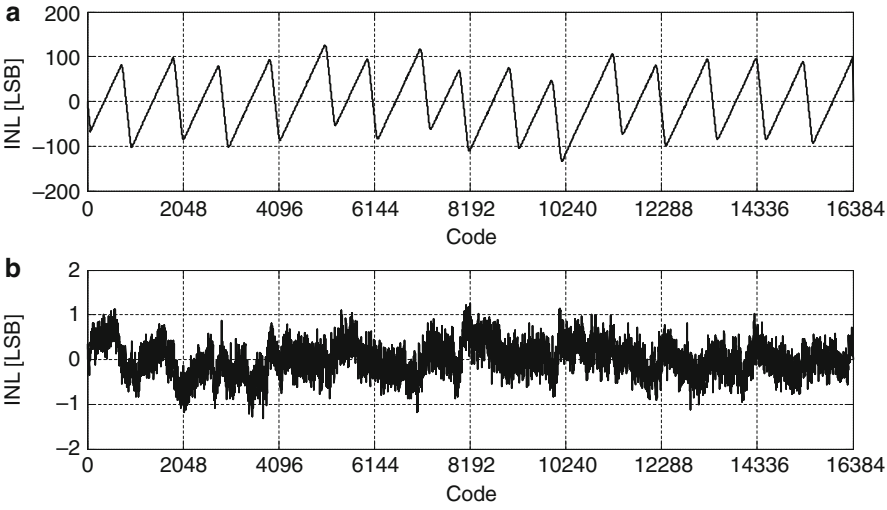


Fig. 3.9 Measured INL. **(a)** Before calibration. **(b)** After calibration

expected in the design phase. This can be attributed to underestimated parasitics and process characterization data stemming from an early phase of process development. The calibration however, next to the benefits discussed above, makes the converter immune to these design inaccuracies. After calibration the DNL is 0.9 LSB, and the INL is 1.3 LSB.

Figure 3.10a shows the measured output spectrum at a 21 MHz input frequency and a 100 MS/s sampling rate. The measured SNR and SFDR are 73 dB and 90 dB

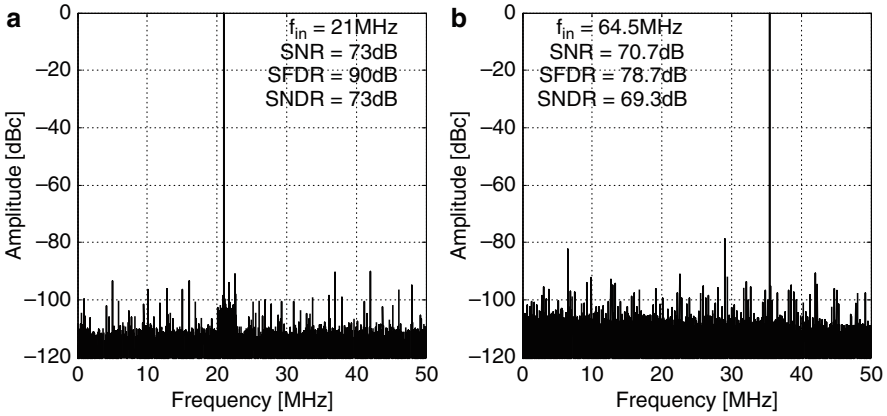


Fig. 3.10 Measured output spectrum at $f_s = 100\,MS/s$. **(a)** $f_{in} = 21\,MHz$. **(b)** $f_{in} = 64.5\,MHz$

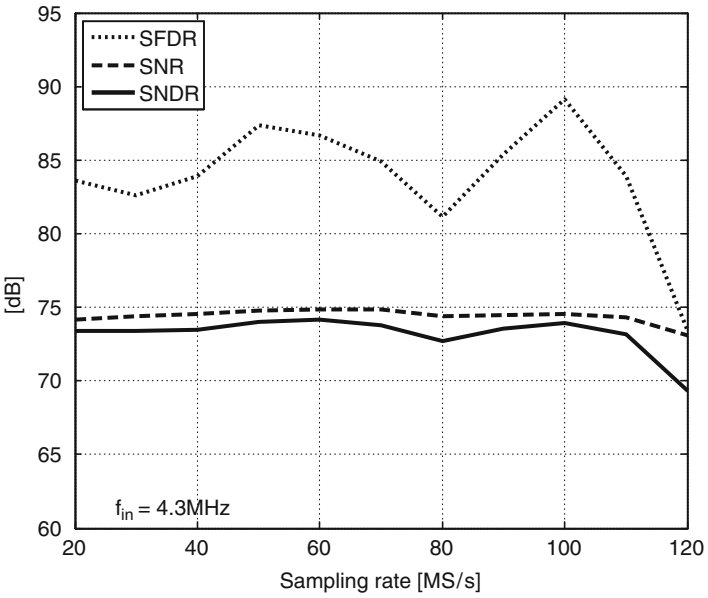


Fig. 3.11 Dynamic performance versus sampling rate

respectively, and the SNDR is 73 dB. Note that the noise floor around the carrier is higher due to the band-pass filtered noise contribution of the signal generator. The output spectrum at a 64.5 MHz input frequency is shown in Fig. 3.10b. It can be seen that the SFDR has decreased to 78.7 dB due to second-order harmonic distortion. The third-order harmonic distortion is at −82 dBc and all other spurious tones are below −90 dBc.

The dynamic performance versus sampling rate is shown in Fig. 3.11, measured at a 4.3 MHz input frequency. The SNR is higher than 73 dB up to a 120 MS/s

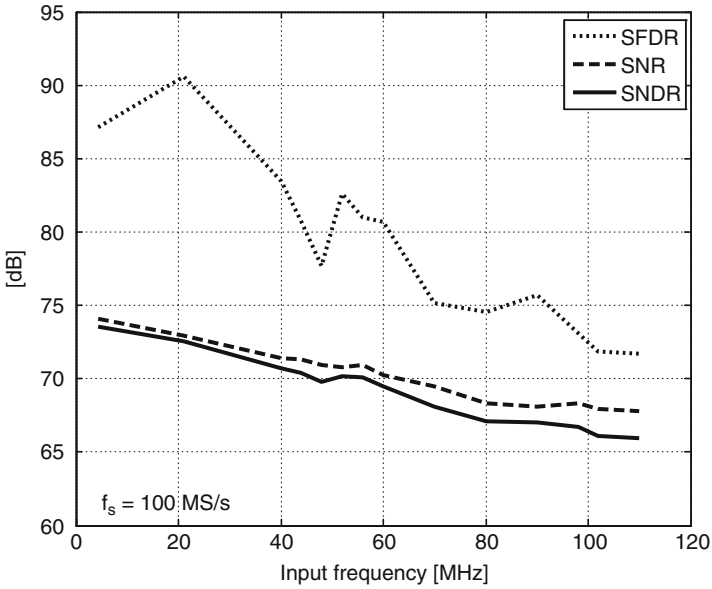


Fig. 3.12 Dynamic performance versus input frequency

sampling rate, and the SFDR remains higher than 81 dB up to a 110 MS/s sampling rate. The dip in SFDR at 80 MS/s is attributed to a test-board issue.

Figure 3.12 shows the dynamic performance versus input frequency at a 100 MS/s sampling rate. The performance degrades gradually towards higher input frequency. The degradation in SNR is due to sampling clock jitter, which is estimated to be 750 fs_{RMS}. The SFDR has a peak value of 90 dB at a 21 MHz input frequency, and rolls off due to second-order harmonic distortion to 72 dB at 110 MHz, while all other spurious tones – including the third-order harmonic distortion – remain below −80 dBc. This roll-off is attributed to an imbalance in the input network on the test-board.

Table 3.1 summarizes the measured performance. Figure 3.13 shows a comparison of this design with published 14 b Nyquist-rate CMOS ADCs. This ADC's figure-of-merit (FOM) is equal to or better than the state-of-the-art at a 3 V supply voltage [1], depending on if the power consumption in the on-chip reference buffers is included or not. It is defined as:

$$FOM = \frac{P}{2^{ENOB} \min(2 \cdot ERBW, f_s)} \tag{3.11}$$

where ENOB is the effective-number-of-bits at low input frequency and ERBW is the effective-resolution-bandwidth.

For all measurements reported above, the clock circuits were separately supplied with 1.8 V. At a 1.2 V clock supply, the bootstrap circuit proved to be ineffective due

Table 3.1 Summary of measured performance

Technology	1-poly 7-metal 90 nm CMOS
Supply voltage	1.2 V
Resolution	14 b
Sampling rate	100 MS/s
Input range	1.6 Vpp
DNL	0.9 LSB
INL	1.3 LSB
SNR	73 dB (fin = 21 MHz), 70.7 dB (fin = 64.5 MHz)
SFDR	90 dB (fin = 21 MHz), 78.7 dB (fin = 64.5 MHz)
SNDR	73 dB (fin = 21 MHz), 69.3 dB (fin = 64.5 MHz)
Power	250 mW
	200 mW (excluding on-chip reference buffers)
FOM	0.68 pJ/conv
	0.55 pJ/conv (excluding on-chip reference buffers)

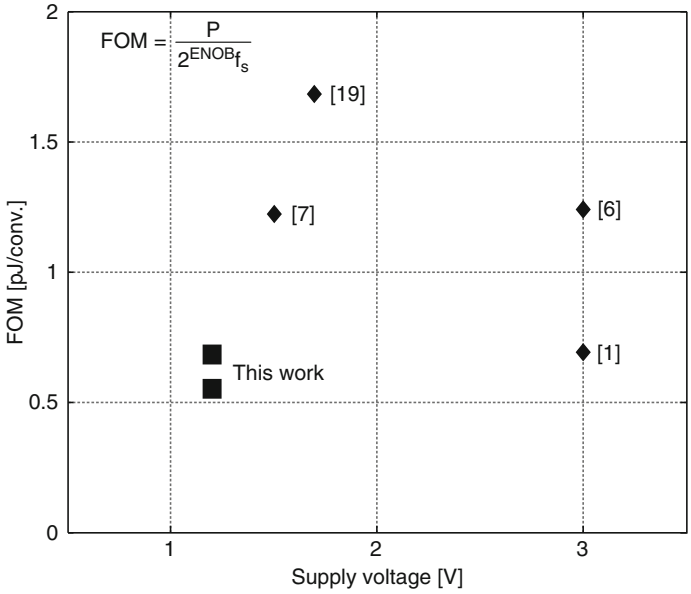


Fig. 3.13 FOM versus supply voltage for 14 b Nyquist-rate CMOS ADCs

to an underestimated parasitic capacitance of the bootstrapped clock line, causing a larger than expected capacitive voltage division of the bootstrapping voltage. The 5 mW increase in power consumption of the clock circuits is included in the overall 250 mW power consumption.

3.6 Conclusions

High-resolution wide-bandwidth ADCs in nm-CMOS allow significant cost reductions in wireless infrastructure systems. This paper discussed smart techniques that enable achieving high linearity and good noise performance in nm-CMOS with a low supply voltage.

Digital calibration of DAC and stage gain non-linearity in a switched-capacitor pipeline ADC enables high linearity, while taking advantage of the excellent digital capabilities of nm-CMOS. For good noise performance the input voltage swing is maximized with a range-scaling technique in the first pipeline stage. The design constraints of an SHA-less architecture, enabling significant power savings, are discussed. A charge-reset switch eliminates the ISI-induced distortion mechanism.

A 14 b 100 MS/s pipeline ADC has been presented that is implemented in 90 nm CMOS technology and operates with a 1.2 V supply voltage. The state-of-the-art power efficiency proves that for this type of converters, low power consumption is feasible in nm-CMOS with a low supply voltage.

Acknowledgments The author would like to thank Berry Buter, Maarten Vertregt, Gerard van der Weide, Govert Geelen, Edward Paulus and Hendrik van der Ploeg for contributions to this work, and Joost Briaire, Kostas Doris, Pieter van Beek, Marcel Pelgrom and other members of the High-Speed Data Converter cluster for many fruitful discussions.

References

1. B.-G. Lee et al., "A 14 b 100 MS/s Pipelined ADC with a Merged Active S/H and First MDAC", *ISSCC Dig. Tech. Papers*, pp. 248–249, Feb. 2008.
2. A. van Roermund et al., "Smart AD and DA Converters", *Proc. ISCAS*, pp. 4062–4065, May 2005.
3. H. Van de Vel, B. Buter, H. van der Ploeg, M. Vertregt, G. Geelen and E. Paulus, "A 1.2 V 250 mW 14b 100MS/s Digitally Calibrated Pipeline ADC in 90 nm CMOS", *VLSI Circuits Symp. Dig*., pp. 74–75, Jun. 2008.
4. H. Van de Vel, B.A.J. Buter, H. van der Ploeg, M. Vertregt, G.J.G.M. Geelen and E.J.F. Paulus, "A 1.2-V 250-mW 14-b 100-MS/s Digitally Calibrated Pipeline ADC in 90-nm CMOS", *IEEE J. Solid-State Circuits*, vol. 44, pp. 1047–1056, Apr. 2009.
5. S.H. Lewis and P.R. Gray, "A Pipelined 5-Msample/s 9-bit Analog-to-Digital Converter", *IEEE J. Solid-State Circuits*, vol. SC-22, pp. 954–961, Dec. 1987.
6. W. Yang et al., "A 3-V 340-mW 14-b 75-Msample/s CMOS ADC with 85-dB SFDR at Nyquist Input", *IEEE J. Solid-State Circuits*, vol. 36, pp. 1931–1936, Dec. 2001.
7. P. Bogner et al., "A 14 b 100 MS/s Digitally Self-Calibrated Pipelined ADC in 0.13 μm CMOS", *ISSCC Dig. Tech. Papers*, pp. 832–833, Feb. 2006.
8. B. Murmann and B.E. Boser, "A 12-bit 75-MS/s Pipelined ADC Using Open-Loop Residue Amplification", *IEEE J. Solid-State Circuits*, vol. 38, pp. 2040–2050, Dec. 2003.
9. J.K. Fiorenza et al., "Comparator-Based Switched-Capacitor Circuits for Scaled CMOS Technologies", *IEEE J. Solid-State Circuits*, vol. 41, pp. 2658–2668, Dec. 2006.
10. M. Anthony, E. Kohler, J. Kurtze, L. Kushner and G. Sollner, "A Process-Scalable Low-Power Charge-Domain 13-bit Pipeline ADC", *VLSI Circuits Symp. Dig*., pp. 222–223, Jun. 2008.
11. B. Murmann, "A/D Converter Trends: Power Dissipation, Scaling and Digitally Assisted Architectures", *Proc. CICC*, pp. 105–112, Sep. 2008.

12. B.-S. Song, S.-H. Lee and M.F. Tompsett, "A 10-b 15-MHz CMOS Recycling Two-Step A/D Converter", *IEEE J. Solid-State Circuits*, vol. 25, pp. 1328–1338, Dec. 1990.
13. A.N. Karanicolas et al., "A 15-b 1-Msample/s Digitally Self-Calibrated Pipeline ADC", *IEEE J. Solid-State Circuits*, vol. 28, pp. 1207–1215, Dec. 1993.
14. I. Galton, "Digital Cancellation of D/A Converter Noise in Pipelined A/D Converters", *IEEE Trans. Circuits and Systems II*, pp. 185–196, Mar. 2000.
15. P. Wambacq and W. Sansen, "*Distortion Analysis of Analog Integrated Circuits*", Kluwer, 1998.
16. K. Bult, "Analog Design in Deep Sub-Micron CMOS", *Proc. ESSCIRC*, pp. 126–132, Sep. 2000.
17. A.-J. Annema, B. Nauta, R. van Langevelde and H. Tuinhout, "Analog Circuits in Ultra-Deep-Submicron CMOS", *IEEE J. Solid-State Circuits*, vol. 40, pp. 132–143, Jan. 2005.
18. S. Limotyrakis, S.D. Kulchycki, D.K. Su and B.A. Wooley, "A 150-MS/s 8-b 71-mW CMOS Time-Interleaved ADC", *IEEE J. Solid-State Circuits*, vol. 40, pp. 1057–1067, May 2005.
19. Y. Chiu, P.R. Gray and B. Nikolic, "A 14-b 12-MS/s CMOS Pipeline ADC With Over 100-dB SFDR", *IEEE J. Solid-State Circuits*, vol. 39, pp. 2139–2151, Dec. 2004.

Chapter 4
A Signal Processing View on Time-Interleaved ADCS

Christian Vogel

Abstract The idea of time-interleaved ADCs (TI-ADCs) is old, but it took more than 25 years until the requirements on converters and the possibility of advanced digital post correction made this architecture attractive. We investigate time-interleaved ADCs with a focus on the involved signal processing. By establishing a discrete-time model of a TI-ADC, we explicitly show that a TI-ADC with mismatches is a time-varying system producing spurious images. This view will help to understand the principles of digital calibration of linear mismatches in TI-ADCs. Currently, time offset mismatches are investigated most extensively. Therefore, we will primarily discuss digital calibration of time offset mismatches, but will generalize the results to frequency response mismatches whenever possible.

4.1 Introduction

The progress in modern electronic systems significantly stems from aggressive downscaling of integrated circuit technologies. Semiconductor companies have made tremendous efforts to keep up with Moore's law and to double the number of transistors per die every 2 years. The performance improvement of digital circuits is a direct result of scaling integrated circuit technologies. Decreasing transistor dimensions and decreasing supply voltages of integrated circuits allow for highly integrated and fast digital operations at low energy levels. Lead microprocessors have shown a doubling in their computing power about every 15 months and a reduction in energy per logic transition of about 65% for each new technology generation [1].

By contrast, analog circuits only partially benefit or even suffer from technology scaling, wherefore a growing gap in the performance of analog and digital circuits becomes observable. This growing performance gap suggests using digital signal processing to improve the inadequacies of analog circuits. The amount of additional digital signal processing depends on the required signal fidelity. Advanced digital methods become possible for moderate to high signal fidelity, e.g., ENOB > 8 b [1].

C. Vogel (✉)
Signal and Information Processing Laboratory, ETH Zurich, CH-8092, Switzerland
e-mail: c.vogel@ieee.org

A.H.M. van Roermund et al. (eds.), *Analog Circuit Design: Smart Data Converters, Filters on Chip, Multimode Transmitters*, DOI 10.1007/978-90-481-3083-2_4,
© Springer Science+Business Media B.V. 2010

Time-interleaved ADCs (TI-ADCs) can take advantage of these technology trends and have recently gained much attention. The principle idea is more than 25 years old [2], but it has taken until now that TI-ADCs with integrated post correction capabilities are used to further push the design limits.

4.2 Time-Interleaved ADCs

As shown in Fig. 4.1, a TI-ADC is a system of M parallel channel ADCs [2]. It takes samples in a time-interleaved way, which is illustrated in Fig. 4.2. While the sampling frequency $f = 1/T$ of the entire TI-ADC fulfils the Nyquist criterion,

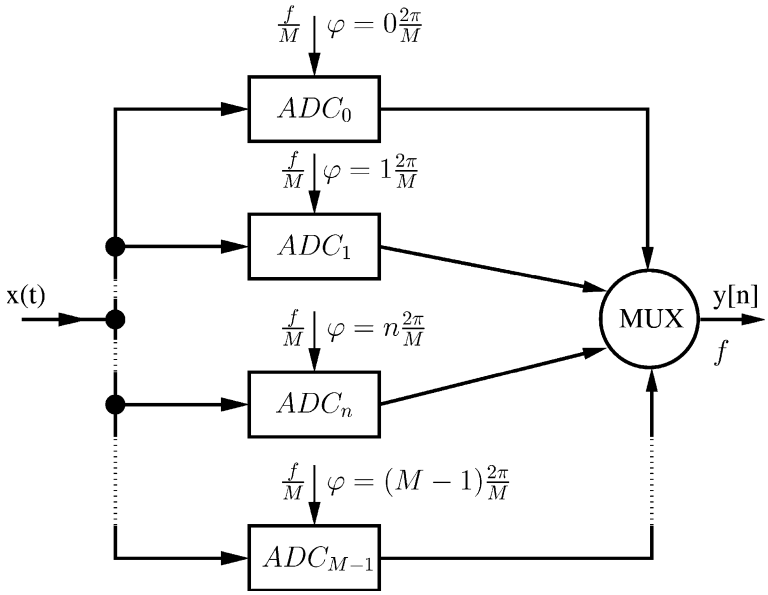


Fig. 4.1 Time-interleaved ADC

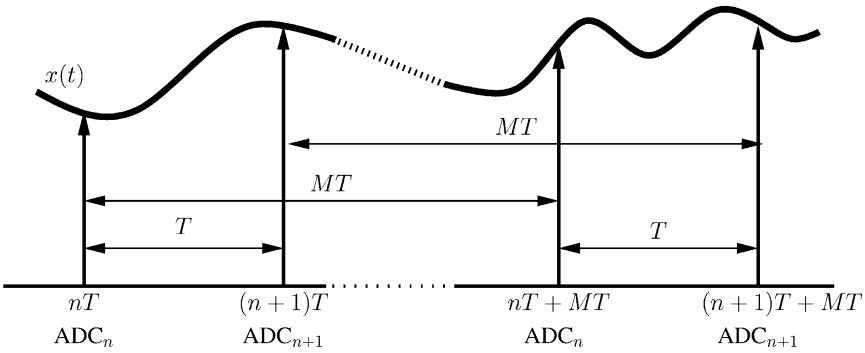


Fig. 4.2 Sampling with an ideal TI-ADC

the sampling frequency $f_{\mathrm{ADC}} = f/M$ of a single channel does not. Ideally, sampling with a TI-ADC with M channels is equivalent to sampling with an ADC at an M times higher sampling rate. Hence, theoretically, we can increase the sampling rate of a TI-ADC by the number of parallel channels. In practice, however, channel mismatches limit the performance of TI-ADCs. Furthermore, each channel has to sample the entire input signal $x(t)$, and, therefore, the sample-and-hold in each channel has to resolve the full input signal bandwidth. In consequence, it is mainly the quantization process that can take advantage from time interleaving. The channels of a TI-ADC can be realized in different converter technologies to achieve for example high-rate and low-power ADCs [3] or high-rate and medium-resolution ADCs [4].

4.3 Modeling Time-Interleaved ADCs

Without considering nonlinear effects, such as the offset, we can represent a TI-ADC as linear M-periodic time-varying system. To show this, we first denote the transfer characteristics of each channel ADC by a linear Filter $H_n(j\Omega)$ for $n = 0, \ldots, M-1$. This linear channel model is shown in Fig. 4.3 and can be expressed in the time domain as

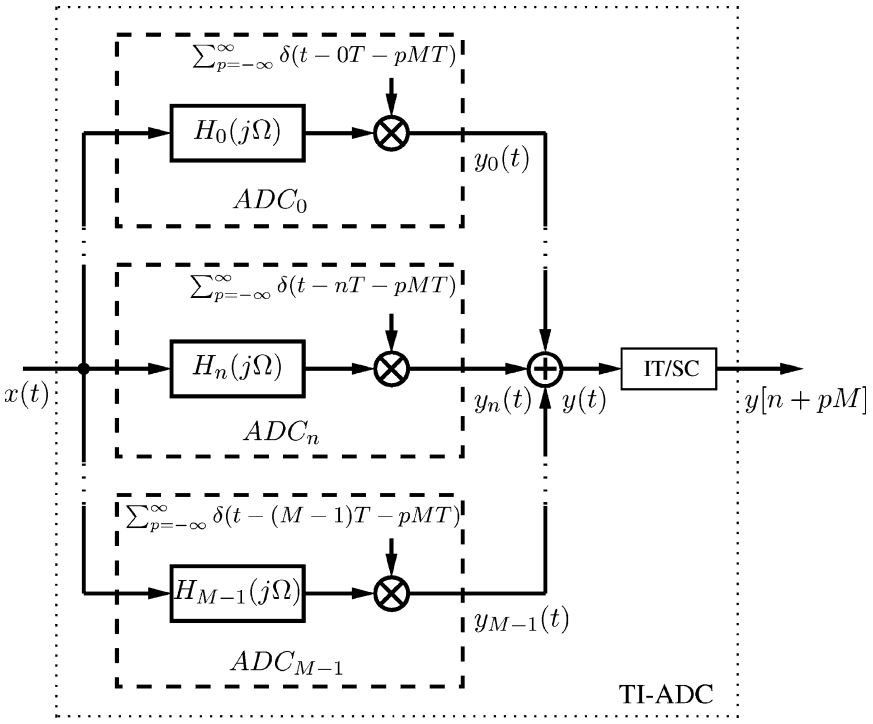


Fig. 4.3 Linear channel model of a TI-ADC

$$y(t) = \sum_{p=-\infty}^{\infty} \sum_{n=0}^{M-1} x(t) * h_n(t)\, \delta(t - nT - pMT) \tag{4.1}$$

and in the frequency domain as [4]

$$Y(j\Omega) = \sum_{p=-\infty}^{\infty} \sum_{k=0}^{M-1} X\left(j\left(\Omega - k\frac{2\pi}{MT} - p\frac{2\pi}{T}\right)\right) \breve{H}_k\left(j\left(\Omega - k\frac{2\pi}{MT} - p\frac{2\pi}{T}\right)\right) \tag{4.2}$$

with

$$\breve{H}_k(j\Omega) = \frac{1}{M} \sum_{n=0}^{M-1} H_n(j\Omega)\, e^{-jkn\frac{2\pi}{M}} \tag{4.3}$$

For each sampling instant $n + pM$, a different channel with corresponding frequency response $H_n(j\Omega)$ is active and processes the sample. If all frequency responses $H_n(j\Omega)$ are identical, there is no difference to a single channel ADC. As soon as the TI-ADC comprises mismatches among the channels, we have, in consequence, for each sample instant $n + pM$ a different frequency response and hence, a time-varying system. Moreover, predetermined by the structure, after M sampling instants, the same channel, i.e., the same frequency response, is active again and therefore we have an M-periodic time-varying system. Since this leads to an M-periodic time-varying system we see modulated images of the input signal in the output.

Traditionally, we distinguish between gain, time offset, and frequency response mismatches. In principal they are all included in the frequency responses $H_n(j\Omega)$, but can be explicitly written as

$$H_n(j\Omega) = g_n S_n(j\Omega)\, e^{-j\Omega\Delta_n T} \tag{4.4}$$

where g_n are the gains leading to gain mismatches, $\Delta_n T$ are the time offsets leading to linear-phase mismatches, and $S_n(j\Omega)$ are the remaining frequency responses leading to frequency response mismatches. Gain and time offset mismatches have attracted the most attention, since they typically have a significant impact on the TI-ADC performance and are easy to correct. To further push the matching limits of TI-ADCs, however, we also have to compensate for frequency response mismatches. Furthermore, by analyzing them, one can gain much more insight on the impact of mismatches in general. By expressing the TI-ADC model in discrete-time, we can further investigate the impact of these modulation images and can develop digital compensation concepts.

Assuming a bandlimited input signal $x(t)$, i.e.,

$$X(j\Omega) = 0 \text{ for } |\Omega| \geq \frac{\pi}{T} \tag{4.5}$$

we can express in discrete time the input signal as

$$X\left(e^{j\omega}\right) = \frac{1}{T} X\left(\frac{\omega}{T}\right) \text{ for } |\omega| < \frac{\pi}{T} \tag{4.6}$$

and the channel frequency responses as

$$H_n\left(e^{j\omega}\right) = H_{n \bmod M}\left(\frac{\omega}{T}\right) \text{ for } n \in \mathrm{Z} \text{ and } |\omega| < \frac{\pi}{T} \tag{4.7}$$

Therefore, we can model a TI-ADC as discrete-time filter bank with modulators

$$Y\left(e^{j\omega}\right) = \sum_{k=0}^{M-1} X\left(e^{j\left(\omega - k\frac{2\pi}{M}\right)}\right) \breve{H}_k\left(e^{j\left(\omega - k\frac{2\pi}{M}\right)}\right) \tag{4.8}$$

with

$$\breve{H}_k\left(e^{j\omega}\right) = \frac{1}{M} \sum_{n=0}^{M-1} H_n\left(e^{j\omega}\right) e^{-jkn\frac{2\pi}{M}} \tag{4.9}$$

or equivalently as discrete-time time-varying system

$$Y_n\left(e^{j\omega}\right) = X\left(e^{j\omega}\right) H_n\left(e^{j\omega}\right) \tag{4.10}$$

with $H_n\left(e^{j\omega}\right) = H_{n+M}\left(e^{j\omega}\right)$ and the discrete-time output signal

$$y[n] = \frac{1}{2\pi} \int_{-\pi}^{\pi} Y_n\left(e^{j\omega}\right) e^{-j\omega n} \tag{4.11}$$

The discrete-time model of a TI-ADC and the two different interpretations are shown in Fig. 4.4. By considering Eqs. 4.8–4.10 and Fig. 4.4, the true nature of mismatches can be seen. The convolution of the signal $x[n]$ by an M-periodic time-varying filter $H_n\left(e^{j\omega}\right)$ can be expressed as the convolution of $x[n]$ with M parallel time-invariant filters $\breve{H}_k\left(e^{j\omega}\right)$ and the modulation with $e^{jk\frac{2\pi}{M}n}$ afterwards. The M modulated signals are summed up to result in the output $y[n]$. The frequency responses $H_n\left(e^{j\omega}\right)$ and $\breve{H}_k\left(e^{j\omega}\right)$ are related by a discrete-time Fourier series given in Eq. 4.9. The output signal and the intermediate signals of Fig. 4.4 are illustrated in Fig. 4.5 for the two-channel case. The input signal filtered by $\breve{H}_0\left(e^{j\omega}\right)$ is not modulated and is comparable to the output of a single channel ADC. All other signal components filtered by $\breve{H}_k\left(e^{j\omega}\right)$ and modulated by $e^{jk\frac{2\pi}{M}n}$ are unwanted images of the input signal degrading the performance of the TI-ADC. Therefore, to digitally compensate the mismatches, the modulated signal images have to be attenuated.

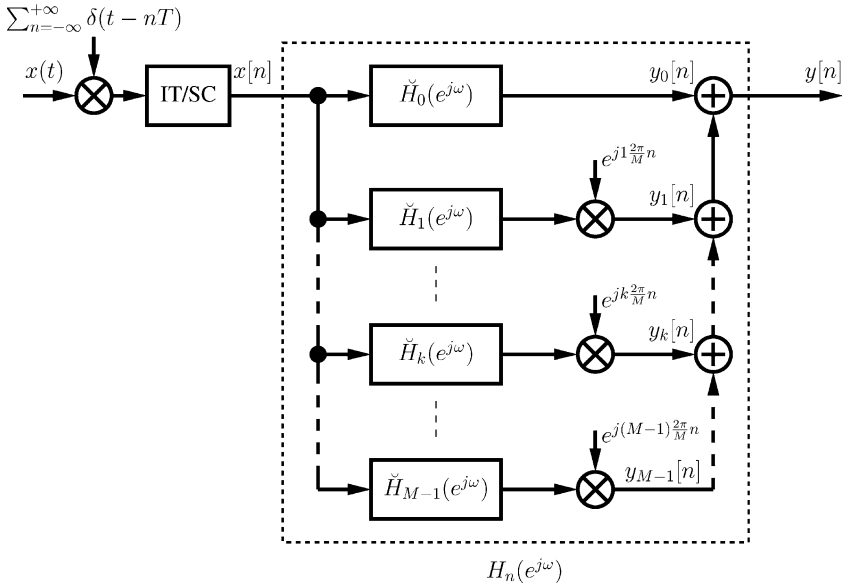


Fig. 4.4 Discrete-time model of a TI-ADC

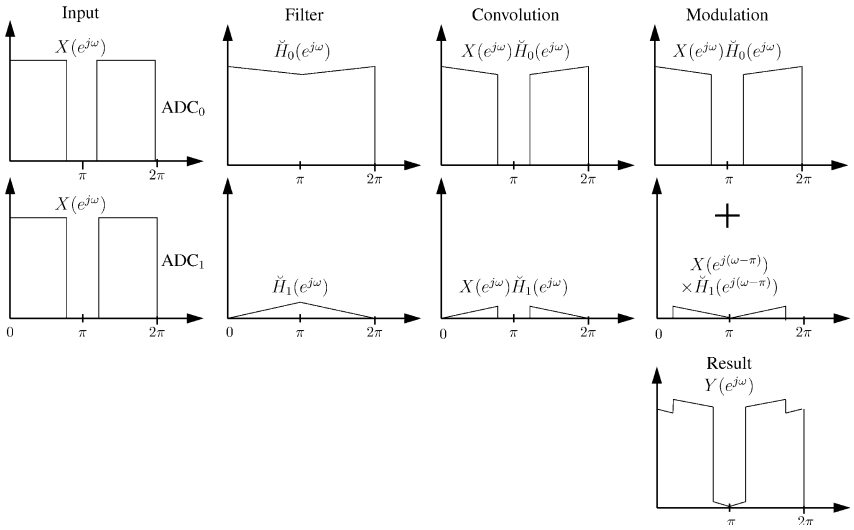


Fig. 4.5 Illustration of the modulation effect due to mismatches ($M = 2$)

4.4 Digital Calibration of Linear Channel Mismatches

The main concern in designing TI-ADCs is to avoid mismatches. Continuing downscaling of feature sizes in integrated circuits and increasing clock rates make the design task even more challenging. Moreover, the matching is subject to time-varying

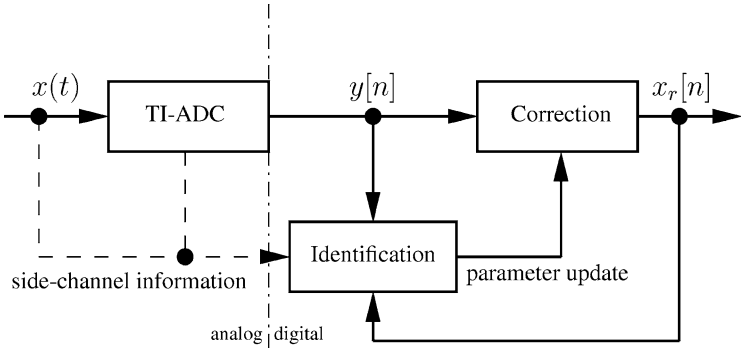


Fig. 4.6 Digital calibration of channel mismatches in TI-ADCs

parameters such as temperature or component aging and drift over time. Therefore, calibration methods, which tune the component matching, e.g., [6, 7], or digitally correct the distorted output signal, e.g., [8, 9], have been proposed. We focus on the digital calibration of channel mismatches. The principal calibration process is illustrated in Fig. 4.6.

The TI-ADC converts the analog input signal $x(t)$ into the digital output signal $y[n]$, which suffers from spurious images due to mismatches in the TI-ADC. In order to reduce these images, the mismatches have to be identified. For this purpose, the identification can use the output signal $y[n]$, the corrected signal $x_r[n]$, and possible side-channel information. Side-channel information includes all the possible knowledge about the input signal, the input signal statistics, environment parameters, or internal signals of the TI-ADC. This knowledge could simplify the identification task, but we either need more information about the application or we have to use additional sensors. In many cases, however, the only assumptions are bandlimitation and some basic statistics of the input signal. Hence, we have to rely on blind identification methods. Unfortunately, it is much more challenging to obtain reliable estimates for the mismatch parameters with blind identification methods. After the mismatch parameters have been identified, the spurious images of the output signal can be digitally removed. First, we are going to discuss digital correction methods and then review digital identification methods.

4.4.1 Digital Correction Methods

Mismatches introduce modulated images of the input signal and it is the goal of digital correction to attenuate these images as good as possible. For the investigation of different correction methods, we assume that the mismatches are known. We will investigate the correction of gain mismatches, linear-phase mismatches (time offset mismatches), and frequency response mismatches. Digital correction is a trade-off between design costs and implementation costs. For example, in order to digitally

correct mismatches with an FIR filter, we have to derive the coefficients of the filter from a set of specifications and measured parameters such as the time offsets. Design, i.e., finding the filter coefficients, and implementation cause costs in terms of power consumption and chip area. The best trade-off depends on the application. In a digital oscilloscope, for example, where power consumption and additional calibration cycles are not a concern, implementation costs, i.e., minimum latency, accuracy, speed-constraints, are much more important than design costs. By contrast, in a receiver for mobile phones with fast changing environments, design costs could become more important than implementation costs. Hence, we need efficient and flexible correction structures. In this context an efficient structure means to have a minimum of additional digital elements, i.e., adders, multipliers, memory, and digital control logic, for a given specification of the correction quality. A flexible structure can easily follow changes of the channel mismatches due to environmental changes such as temperature and aging. A general multiplier can accomplish the digital correction of gain mismatches. Time-offset mismatches and frequency response mismatches need to be discussed in more detail.

4.4.1.1 Time Offset Mismatches

Time-offset mismatches Δ_n are caused by different signal delays among the clock paths and the channel paths. This is illustrated in Fig. 4.7.

The channel delays the input signal $x(t)$ by Δ_n^{CH}, i.e., $x\left(t-\Delta_n^{\mathrm{CH}}\right)$ for the nth channel, and is sampled at time $nT+\Delta_n^{\mathrm{CLK}}$ resulting in an overall delay of $\Delta_n=\Delta_n^{\mathrm{CH}}-\Delta_n^{\mathrm{CLK}}$. For pure digital correction the origin of the time offsets is not important, but in the analog domain we can achieve matching by either employing analog delay elements in the clock path or by changing the delay of the signal path. As mentioned earlier, because of mismatches a TI-ADC is an M-periodic time-varying system. Therefore, a discrete-time time-varying filter $Q_n\left(e^{j\omega}\right)$ can compensate the effect as shown in Fig. 4.8. A discrete-time time-varying filter changes its frequency response for each time instant n. By using a time-varying finite impulse response (FIR) filter given by

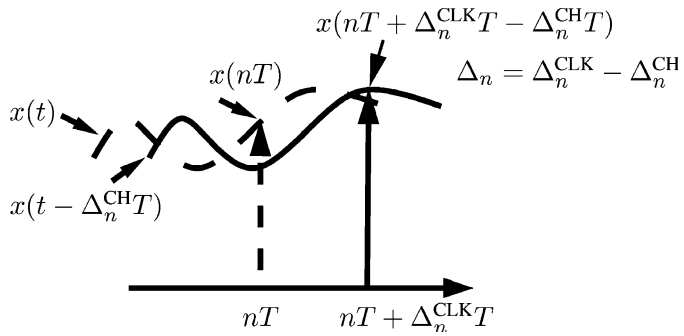


Fig. 4.7 Principle error sources of time offset mismatches

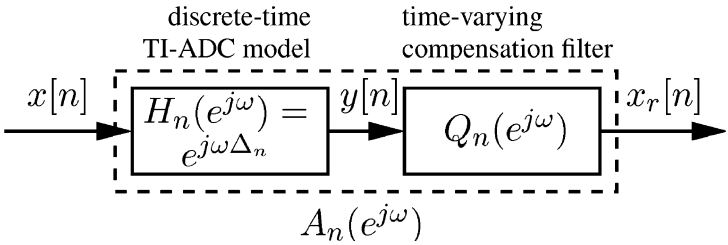


Fig. 4.8 Correction of time offsets through a discrete-time time-varying filter

$$Q_n\left(e^{j\omega}\right) = \sum_{k=0}^{K} q_n\left[k\right] e^{-j\omega k} \tag{4.12}$$

where K is the order of the filter, the frequency response of the cascaded system results in [10]

$$A_n\left(e^{j\omega}\right) = \sum_{k=0}^{K} e^{j\omega\Delta_{n-k}} q_n\left[k\right] e^{-j\omega k} \tag{4.13}$$

By optimizing the filter coefficients of $q_n\left[k\right]$ we can control the overall frequency $A_n\left(e^{j\omega}\right)$. Ideally, the frequency response $A_n\left(e^{j\omega}\right)$ should have a unity frequency response over the frequency range $-\pi \leq \omega < \pi$. In practice, however, it is not possible to find such coefficients. Therefore, the optimization problem is relaxed by allowing some additional delay, often $K/2$, and introducing some "don't care" band. The optimization problem to find the filter coefficients in the L_p-norm sense can then be written as

$$\arg \min_{q_n \in R^{K+1}} \left\| A_n\left(e^{j\omega}\right) - e^{-j\omega\frac{K}{2}} \right\|_p \text{ for } |\omega| < \omega_c < \pi \text{ and } n = 0, \ldots, M-1 \tag{4.14}$$

where typically L_2 or L_∞ norms are used. Hence, for each channel with time offset Δ_n we create a coefficient set $q_n = [q_n\left[0\right], q_n\left[1\right], \ldots, q_n\left[K\right]]$. For each time instant n the FIR filter is updated with the corresponding coefficient set. With this approach the optimal filter coefficients can be found, but the optimization procedure has to be repeated each time a time offset changes. As initially discussed, this can become impossible for on chip solutions. In principle, by using multivariate polynomial filter structures, the filter design can be done entirely off-line [11]. Unfortunately, even for a low number of channels and a low number of filter coefficients, the problem becomes quickly computational unfeasibly [11]. Therefore we have to find a different structure, which simplifies the optimization process and still achieves an acceptable performance. The error reconstruction principle is a possible solution for the flexible digital correction of the time offset problem [12, 13]. It is illustrated in Fig. 4.9.

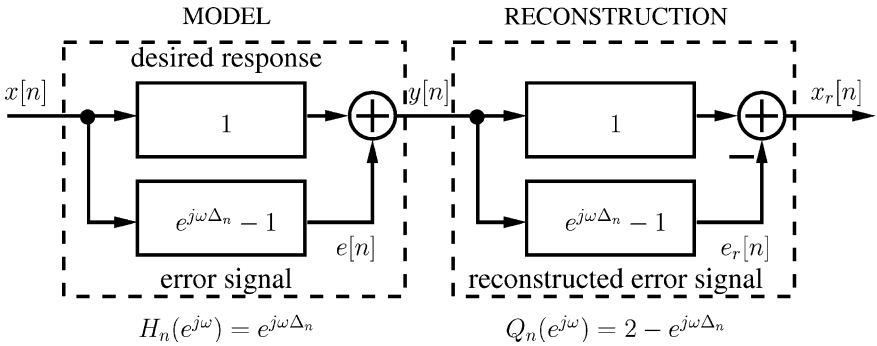


Fig. 4.9 Reconstruction of time offset mismatches through the error reconstruction principle

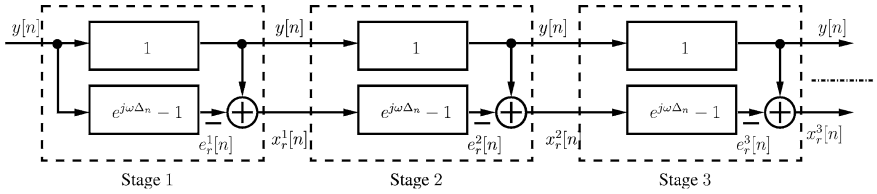


Fig. 4.10 Cascading the error reconstruction principle

On the left side, we have again the discrete-time model of a TI-ADC consisting of a time-varying filter $H_n\left(e^{j\omega}\right)$. In contrast to Fig. 4.8, we split the frequency response $H_n\left(e^{j\omega}\right)$ into two responses: the desired response, i.e., 1, and the error response consisting of the modulated images, i.e., $e^{j\omega\Delta_n} - 1$. Since we have a linear system, the overall frequency response is the linear-phase response $H_n\left(e^{j\omega}\right) = e^{j\omega\Delta_n}$, but the separation helps to explain the reconstruction principle. Using the output signal $y\,[n]$ of the TI-ADC, we reconstruct the error by using the same time-varying filter as in the model. The reconstructed error signal $e_r\,[n]$ is then subtracted from the output signal. The reconstructed error signal $e_r\,[n]$ is different from the error signal $e\,[n]$, because the input signal is $y\,[n]$ and not $x\,[n]$. For typical timing offsets the energy of the error signal $e\,[n]$ is small compared to $x\,[n]$, and the reconstruction error signal $e_r\,[n]$ is close to $e\,[n]$ leading to a reconstructed signal $x_r\,[n]$, which is closer in the L_2 sense to the ideal signal $x\,[n]$ than $y\,[n]$. For typical examples, the improvement of the SNR is about two times the SNR before correction. As the output signal $x_r\,[n]$ is closer to $x\,[n]$, repeated reconstruction with the same filter structure as shown in Fig. 4.10, progressively improves the SNR.

This structure does not only work for TI-ADCs, but can be applied for all nonuniformly sampled signals with time offsets that are small compared to the sampling period. On the one hand, compared to the optimal filter design approach, the structure needs more adders and multipliers to achieve the same SNR. On the other hand, the filter design complexity can be significantly reduced, since the reconstruction filter is determined by the structure and only depends on the current time offset Δ_n

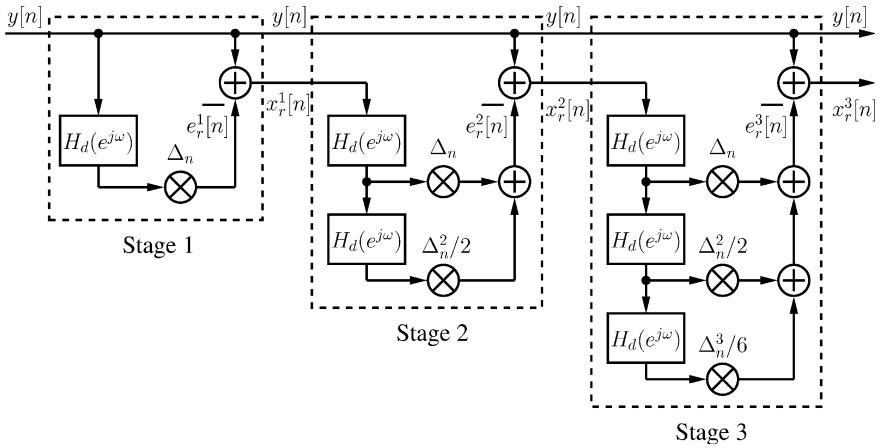


Fig. 4.11 The differentiator-multiplier cascade efficiently implements the error reconstruction principle

and not on past time offsets as in Eqs. 4.12–4.14. Therefore, it is possible to design the reconstruction filters independent from the time offsets. One efficient implementation – the differentiator-multiplier cascade [11] – is shown in Fig. 4.11, where

$$H_d\left(e^{j\omega}\right) = j\omega \quad \text{for } \pi \leq \omega < \pi \tag{4.15}$$

is the frequency response of an ideal discrete-time differentiator.

The structure is obtained by using a Taylor series of the frequency response $e^{j\omega\Delta_n} - 1$, i.e.,

$$\begin{aligned} T\left\{e^{j\omega\Delta_n} - 1\right\} &= j\omega\Delta_n + (j\omega)^2\,{\Delta_n}^2/2 + (j\omega)^3\,{\Delta_n}^3/6 + \ldots \\ &= H_d\left(e^{j\omega}\right)\Delta_n + H_d\left(e^{j\omega}\right)^2\,{\Delta_n}^2/2 + H_d\left(e^{j\omega}\right)^3\,{\Delta_n}^3/6 + \ldots \end{aligned} \tag{4.16}$$

For each reconstruction stage more terms of the Taylor series are used, since the requirements on the reconstruction quality increase and likewise the required approximation accuracy of the frequency response $e^{j\omega\Delta_n} - 1$. The actual frequency response of the reconstruction system is directly determined by the current time offset Δ_n. Therefore, the filter design process is inherently part of the reconstruction structure and is reduced to adjust some general multipliers. For the purpose of a TI-ADC, one stage of the differentiator-multiplier cascade is typically sufficient.

4.4.1.2 Frequency Response Mismatches

In this section, we generalize the results from the last section. Time offset mismatches are linear-phase mismatches in the frequency domain and therefore a

special case of frequency response mismatches. Hence, we can use the same approaches as in the last section with some slight modifications. Without explicitly assuming time offset mismatches, the overall frequency response can be written as [14]

$$A_n\left(e^{j\omega}\right) = \sum_{k=0}^{K} H_{n-k}\left(e^{j\omega}\right) q_n[k]\, e^{-j\omega k} \tag{4.17}$$

where the special case of time offset mismatches is obtained for $H_n\left(e^{j\omega}\right) = e^{j\omega\Delta_n}$, i.e., Eq. 4.13. Accordingly, the same optimization procedure as in Eq. 4.14 can be used, but also the drawback is the same. Each time the frequency responses change due to environmental changes such as temperature or aging, the reconstruction filters have to be redesigned through the costly optimization process. To avoid the complexity of a filter redesign through optimization, we can generalize the error reconstruction principle for time offset mismatches in.

The error reconstruction principle for frequency response mismatches is shown in Fig. 4.12 [5]. The discrete-time frequency response is split into a time-invariant frequency response $\overline{H}\left(e^{j\omega}\right)$ representing the desired signal and a time-varying frequency response $G_n\left(e^{j\omega}\right) = H_n\left(e^{j\omega}\right) - \overline{H}\left(e^{j\omega}\right)$ representing the error signal $e[n]$, i.e., the modulated images. The reconstruction uses the output signal $y[n]$ to produce the reconstructed error signal $e_r[n]$ that is subtracted from the TI-ADC output signal $y[n]$ to remove the error signal $e[n]$. For $\overline{H}_n\left(e^{j\omega}\right) = e^{H_d\left(e^{j\omega}\right)\Delta_n}$ and $\overline{H}\left(e^{j\omega}\right) = 1$, the structure reduces to the one for time offset compensation in Fig. 4.9. The frequency responses of the compensation structure are directly determined by the frequency responses of the channels. This can significantly simplify the redesign complexity. A simple example is the compensation of first-order bandwidth mismatches. Assuming the frequency response mismatches are given by a first order low-pass filter, it can be modeled in discrete-time as

$$H_n\left(e^{j\omega}\right) = \frac{1}{1 + j\,\frac{\omega}{\omega_c \beta_n}} \tag{4.18}$$

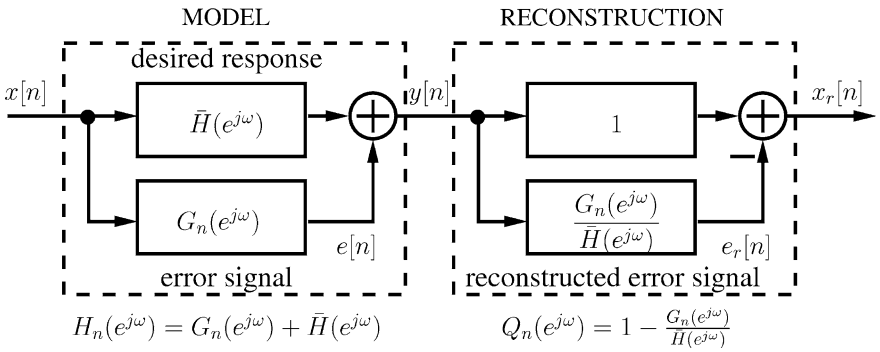


Fig. 4.12 The error reconstruction principle for frequency response mismatches

Thus, the frequency response of the channel $H_n\left(e^{j\omega}\right)$ is determined by the single parameter β_n, which depends on the nth channel. By explicitly expressing the dependency on the parameter, i.e.,

$$H\left(e^{j\omega}, \beta\right) = \frac{1}{1 + j\,\frac{\omega}{\omega_c \beta}} \tag{4.19}$$

we can express the channel frequency response as $H_n\left(e^{j\omega}\right) = H\left(e^{j\omega}, \beta_n\right)$. Setting the desired frequency response to

$$\overline{H}\left(e^{j\omega}\right) = \frac{1}{1 + j\,\frac{\omega}{\omega_c}} \tag{4.20}$$

the compensation filter is given by

$$Q\left(e^{j\omega}, \beta\right) = 1 - \frac{H_n\left(e^{j\omega}, \beta\right) - \overline{H}\left(e^{j\omega}\right)}{\overline{H}\left(e^{j\omega}\right)} \tag{4.21}$$

with $Q_n\left(e^{j\omega}\right) = Q\left(e^{j\omega}, \beta_n\right)$. A Farrow filter [15] can efficiently approximate the frequency response given in Eq. 4.21 and is given by

$$Q^a\left(e^{j\omega}, \beta\right) = 1 - \sum_{l=0}^{L} \sum_{p=1}^{P} c_p[l] \beta^p e^{-j\omega l} \tag{4.22}$$

where L is the order of the subfilters and P is the order of the polynomial. The reconstruction structure for first-order bandwidth mismatches is shown in Fig. 4.13.

The shown example consists of three fixed filters and three general multipliers. The parameter β_n corresponds to the bandwidth mismatch of the nth channel. Hence, for each sample, a new parameter β_n is set and a new frequency response is generated to compensate for the bandwidth mismatches. Therefore, any changes of the bandwidth mismatches and the corresponding parameter β_n due to for example temperature variations, can be easily compensated by the structure and the updated parameters. To improve the reconstruction accuracy, we can also cascade the structures as shown in Fig. 4.10. Nevertheless, for the correction the parameter β_n has to be accurately estimated. Identification methods are discussed in the next section.

4.4.2 Digital Identification Methods

The identification of channel mismatches is the most important task in the channel mismatch calibration process. If the identified parameters are imprecise or even wrong, the most sophisticated correction method cannot improve the quality of the TI-ADC output signal. Basically, we can distinguish between methods with special

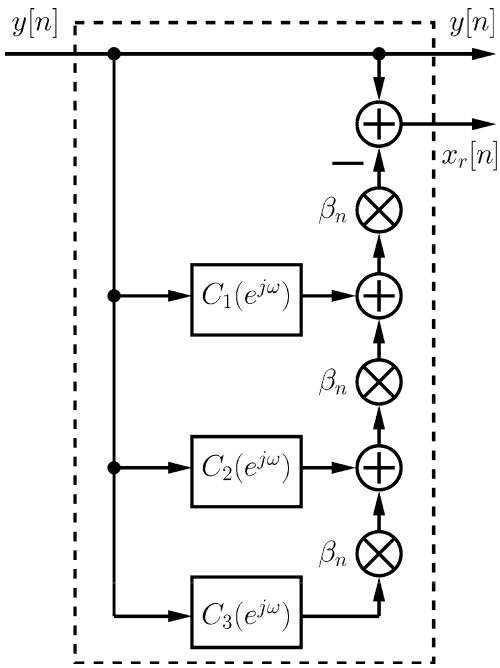


Fig. 4.13 Correction of first-order bandwidth mismatches ($P = 3$)

input signals, i.e., offline methods, and blind methods, i.e., online methods. In many cases, however, the boundaries get blurred, and we have a combination of both approaches.

4.4.2.1 Off-line Identification

We can find accurate solutions for the identification of the channel mismatches with special input signals [16,17]. All of them are based on the same principle. Assuming a TI-ADC with M-channels, the frequency response mismatches can be identified by using a signal bandlimited to $f_s/(2M)$. For such input signals, the input signal and the modulated images do not overlap, which can be realized from Figs. 4.4 and 4.5. Parameters such as gain, timing offset, phase, or amplitude can be derived from the output spectrum. The best identification accuracy can be achieved with a coherently sampled sinusoidal input signal. Applying a sinusoidal input signal $x(t) = \cos(\Omega_0 t)$ to the TI-ADC, we obtain with Eqs. 4.6 and 4.8,

$$Y\left(e^{j\omega}\right) = \sum_{k=0}^{M-1} \pi \left[\delta\left(\omega - \omega_0 - k\frac{2\pi}{M}\right) + \delta\left(\omega + \omega_0 - k\frac{2\pi}{M}\right)\right] \breve{H}_k\left(e^{j\left(\omega - k\frac{2\pi}{M}\right)}\right) \tag{4.23}$$

The TI-ADC output spectrum consists of the spectrum of the sinusoidal input signal filtered by $\breve{H}_0\left(e^{j\omega}\right)$ and of $M-1$ modulated and by $\breve{H}_k\left(e^{j\omega}\right)$ filtered spectral

images. As long as the input frequency ω_0 differs from $k2\pi/M$ the images are mutually spectrally separated. Therefore, we can identify the relative modulated frequency responses at frequency ω_0 by relating the output spectrum as

$$\frac{Y\left(e^{j\left(\omega_0+k\frac{2\pi}{M}\right)}\right)}{Y\left(e^{j\omega_0}\right)} = \frac{\breve{H}_k\left(e^{j\omega_0}\right)}{\breve{H}_0\left(e^{j\omega_0}\right)} \quad \text{for } k = 1, 2, \ldots, M-1 \tag{4.24}$$

By repeating this identification for a sufficient number of frequencies ω_0, we can characterize the relative modulated frequency responses of the channels. By applying the inverse DTFS, we obtain relative frequency responses

$$F_n\left(e^{j\omega_0}\right) = \frac{H_n\left(e^{j\omega_0}\right)}{\breve{H}_0\left(e^{j\omega_0}\right)} = \sum_{k=0}^{M-1} \frac{\breve{H}_k\left(e^{j\omega_0}\right)}{\breve{H}_0\left(e^{j\omega_0}\right)} e^{jkn\frac{2\pi}{M}} \tag{4.25}$$

The relative gain mismatches $g_n/\breve{g}_0$ are given by $\left|F_n\left(e^{j\omega_0}\right)\right|$, and the time offset mismatches Δ_n are given by $\arg\left(F_n\left(e^{j\omega_0}\right)\right)/\omega_0$. In both cases, the result is only true when amplitude and phase mismatches do not change over the frequency. If they do so, however, it is better to use some averaging over an appropriate range of frequencies. The best results can be achieved if the weighting reflects the spectral distribution of the input signal during normal operation of the TI-ADC.

4.4.2.2 On-line Identification

We can find several methods for gain mismatch identification, e.g., [7, 8]. Most of them compare the averaged output power among all channels in some way. Although such methods can be vulnerable to an input signal correlation with the switching sequence of the channels, they work well for appropriate input signals, which most communications signals are. The identification of time offset mismatches is much more demanding. Although we only have to estimate a single parameter for each channel, the requirements on the accuracy are high. Furthermore, the only assumption is an oversampled input signal that provides sufficient energy over some amount of time and bandwidth. With these assumptions, we obtain an output spectrum for a two channel TI-ADC as depicted in Fig. 4.14. Since the input signal is oversampled, we obtain a band, where only modulated images are present, i.e., the mismatch band. This mismatch band can be used to identify the parameters for the time offset mismatches.

A possible solution for the two-channel case is shown in Fig. 4.15 [18]. On the left hand side, we basically see the first stage of the differentiator-multiplier cascade shown in Fig. 4.11 with $\Delta_n = (-1)^n \Delta_0$, since $\Delta_0 = -\Delta_1$. It will be used to reconstruct the signal. Additionally, we have two signal paths running into a high-pass filter (HP). In the upper path, the high-pass filter removes all signal energy outside the mismatch band and leaves only a high-pass filtered error signal $e_{\mathrm{HP}}[n]$.

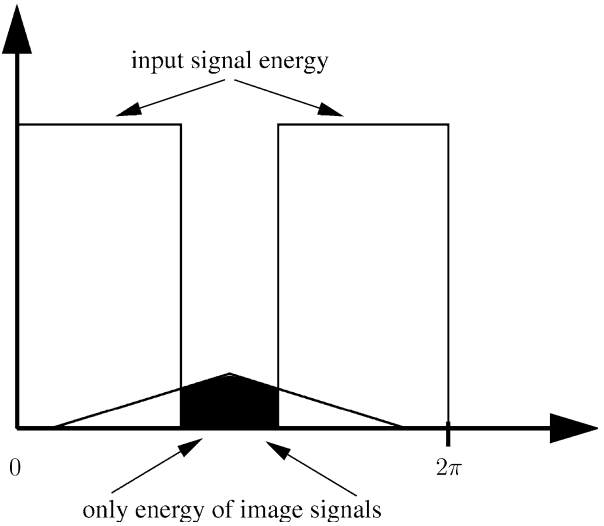


Fig. 4.14 The mismatch band for a four-channel TI-ADC ($M = 2$)

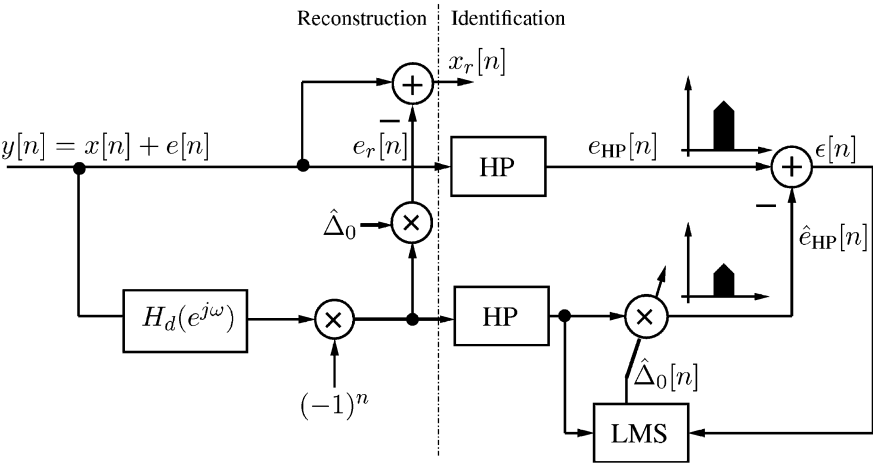


Fig. 4.15 Blind identification of timing mismatches ($M = 2$)

In the lower path, the output signal $y\,[n]$ is filtered and modulated. Multiplied with the correct time offset Δ_0, the modulated signal can be used to obtain the reconstructed signal $x_r\,[n]$. In our setting, however, this parameter is unknown and has to be estimated. Therefore, the modulated signal is high-pass filtered resulting in, as a first approximation, a signal that only differs by the gain Δ_0 from the signal $e_{\mathrm{HP}}\,[n]$. Applying the least-mean square (LMS) algorithm, we can find this gain, which is actually the estimation of the time offset Δ_0. The same principle can be extended to more channels including gain mismatches [19]. Different approaches using multi-rate signal processing are also possible [20, 21].

4.5 Conclusions

We have discussed the digital calibration of mismatches with a particular focus on time offset mismatches. We have very flexible and efficient structures for the digital correction of time offset mismatches, which can basically also be used for frequency response mismatches. Nevertheless, the details of an efficient implementation are still under investigation. The off-line identification of mismatches can be done very accurately. In contrast to that, for on-line identification, i.e., blind identification, we have some methods for the time offset identification, but they are not yet fully reliable.

References

1. B. Murmann, C. Vogel, and H. Koeppl, "Digitally Enhanced Analog Circuits: System Aspects", Proceedings of the 2008 IEEE International Symposium on Circuits and Systems, Seattle (USA), 18–21 May 2008, pp. 560–563.
2. W. C. Black, Jr. and D. A. Hodges, "Time-interleaved Converter Arrays," IEEE Journal of Solid-State Circuits, vol. SC-15, no. 6, Dec. 1980, pp. 1022–1029.
3. D. Draxelmayr, "A 6 b 600 MHz 10 mW ADC Array in Digital 90 nm CMOS," in 2004 IEEE International Solid-State Circuits Conference, vol. 1, Feb. 2004, pp. 45–48.
4. C.-C. Hsu, F.-C. Huang, C.-Y. Shih, C.-C. Huang, Y.-H. Lin, C.-C. Lee, B. Razavi, "An 11 b 800 MS/s Time-Interleaved ADC with Digital Background Calibration", IEEE International Solid-State Circuits Conference, Feb. 2007, pp. 464–615.
5. C. Vogel and S. Mendel, "A Flexible and Scalable Structure to Compensate Frequency Response Mismatches in Time-Interleaved ADCs", IEEE Transactions on Circuits and Systems I: Regular Papers, 2003, pp. 1–1.
6. C. Vogel, D. Draxelmayr, and F. Kuttner, "Compensation of Timing Mismatches in Time-interleaved Analog-to-Digital Converters through Transfer Characteristics Tuning," in Proceedings of the 47th IEEE International Midwest Symposium On Circuits and Systems, vol. 1, Jul. 2004, pp. 341–344.
7. P. J. A. Harpe, J. A. Hegt, A. H. M. van Roermund, "Analog Calibration of Channel Mismatches in Time-Interleaved ADCs", International Journal of Circuit Theory and Applications, vol. 37, no. 2, 2009, pp. 301–318.
8. S. Jamal, D. Fu, M. Singh, P. Hurst, and S. Lewis, "Calibration of Sample-time Error in a Two-channel Time-interleaved Analog-to-Digital Converter," IEEE Transaction on Circuits and Systems I: Regular Papers, vol. 51, no. 1, Jan. 2004, pp. 130–139.
9. C. Vogel, S. Saleem, and S. Mendel, "Adaptive Blind Compensation of Gain and Timing Mismatches in M-channel Time-interleaved ADCs," in Proceedings of the 14th IEEE International Conference on Electronics, Circuits and Systems ICECS, Sep. 2008, pp. 49–52.
10. H. Johansson and P. Löwenborg, "Reconstruction of Nonuniformly Sampled Bandlimited Signals by Means of Time-varying Discrete-time FIR Filters," EURASIP Journal of Applied Signal Processing, vol. 2006, pp. 1–18, 2006, DOI 10.1155/ASP/2006/64185, 64185.
11. H. Johansson, P. Löwenborg, and K. Vengattaramane, "Least-squares and Minimax Design of Polynomial Impulse Response FIR Filters for Reconstruction of Two-periodic Nonuniformly Sampled Signals," IEEE Transaction on Circuits System I, Registered Papers, vol. 54, no. 4, Apr. 2007, pp. 877–888.
12. S. Tertinek and C. Vogel, "Reconstruction of Two-periodic Nonuniformly Sampled Bandlimited Signals Using a Discrete-Time Differentiator and a Time-Varying Multiplier," IEEE Transactions on Circuits and Systems II, vol. 54, no. 7, Jul. 2007, pp. 616–620.

13. S. Tertinek and C. Vogel, "Reconstruction of Nonuniformly Sampled Bandlimited Signals Using a Differentiator-Multiplier Cascade," IEEE Transactions on Circuits and Systems I, vol. 55, no. 8, Sep. 2008, pp. 2273–2286.
14. H. Johansson and P. Lowenborg, "A Least-Squares Filter Design Technique for the Compensation of Frequency Response Mismatch Errors in Time-Interleaved A/D Converters," IEEE Transactions on Circuits and Systems II: Express Briefs, vol. 55, no. 11, Nov. 2008, pp. 1154–1158.
15. C. W. Farrow, "A Continuously Variable Digital Delay Element," in Proceedings of the IEEE International Symposium on Circuits Systems, Espoo, Finland, vol. 3, Jun. 1988, pp. 2641–2645.
16. Y. C. Jenq, "Digital Spectra of Nonuniformly Sampled Signals: a Robust Sampling Time Offset Estimation Algorithm for Ultra High-Speed Waveform Digitizers Using Interleaving," IEEE Transaction on Instrumentation and Measurement, vol. 39, no. 1, Feb. 1990, pp. 71–75.
17. M. Seo, M. Rodwell, and U. Madhow, "Comprehensive Digital Correction of Mismatch Errors for a 400-Msamples/s 80-dB SFDR Time-Interleaved Analog-To-Digital Converter," IEEE Transaction on Microwave Theory and Techniques, vol. 53, no. 3, Apr. 2005, pp. 1072–1082.
18. S. Saleem and C. Vogel, "LMS-Based Identification and Compensation of Timing Mismatches in a Two-Channel Time-Interleaved Analog-to-Digital Converter," Proceedings of the IEEE Norchip Conference 2007, Aalborg (Denmark), Nov. 2007, pp. 19–20.
19. C. Vogel, S. Saleem, and S. Mendel, "Adaptive Blind Compensation of Gain and Timing Mismatches in M-Channel Time-Interleaved ADCs," Proceedings of the 14th IEEE International Conference on Electronics, Circuits and Systems (ICECS 2008), St. Julians (Malta), 1–3 Sep. 2008, pp. 49–52.
20. S. Huang and B.C. Levy, "Blind Calibration of Timing Offsets for Four-Channel Time-Interleaved ADCs," IEEE Transactions on Circuits and Systems I: Regular Papers, vol.54, no.4, Apr. 2007, pp. 863–876.
21. T.-H. Tsai, P. J. Hurst, and S. H. Lewis, "Correction of Mismatches in a Time-Interleaved Analog-to-Digital Converter in an Adaptively Equalized Digital Communication Receiver," IEEE Transactions on Circuits and Systems I: Regular Papers, vol.56, no.2, Feb. 2009, pp. 307–319.

Chapter 5
DAC Correction and Flexibility, Classification, New Methods and Designs

Georgi Radulov, Patrick Quinn, Hans Hegt, and Arthur van Roermund

Abstract This paper classifies correction methods for current-steering Digital-to-Analog Converters (DACs), with an emphasis on self-calibration. Based on this classification, missing methods are identified. Three new DAC correction methods are proposed that can fill in these gaps: high-level mapping, suppression of HD, and calibration of binary currents. All three of them are based on parallel sub-DACs. The paper also proposes to further exploit the advantages of using such parallel sub-DACs to achieve flexibility. Two test-chip implementations in 250 and 180 nm CMOS validate the proposed concepts.

5.1 Introduction

The diversity of the DAC correction methods is very high. That is why this paper proposes a classification which shows the links among the various correction methods, the common properties of them, and missing DAC correction methods. Next to error correction, flexibility is an important issue. For FPGAs to extend into the mixed signal domain, co-integration of flexible high performance ADCs and DACs becomes necessary. Such an approach would successfully address the challenges of modern mixed-signal electronics: time-to-market pressure, increased design complexity, advanced but unreliable CMOS technologies, on the fly tuning to numerous new standards and requirements. The vast and programmable digital resources that are available in the FPGAs can assist the performance, relax the requirements, and improve yield of the co-integrated ADCs and DACs. This paper investigates and classifies known DAC correction methods and identifies missing methods (Sect. 5.2.1); proposes new correction methods, based on parallel

G. Radulov (✉), H. Hegt, and A. van Roermund
Mixed-signal Microelectronics Group, Department of Electrical Engineering, Eindhoven University of Technology, P.O. Box 513, 5600 MB Eindhoven
e-mail: G.Radulov@tue.nl

P. Quinn
Mixed Signal Design Group, Xilinx Ireland, Logic Drive, Citywest Business Campus, Saggart, Co. Dublin, Ireland

A.H.M. van Roermund et al. (eds.), *Analog Circuit Design: Smart Data Converters, Filters on Chip, Multimode Transmitters*, DOI 10.1007/978-90-481-3083-2_5,
© Springer Science+Business Media B.V. 2010

sub-DACs, to fill in these gaps (Sect. 5.2.2); analyzes in detail the self-calibration DAC correction method (Sect. 5.3); and introduces a new flexible DAC architecture (Sect. 5.4). To demonstrate the discussed concepts in practice, two test-chip implementations are presented (Sect. 5.5). Finally, conclusions are drawn.

5.2 Correction Methods for Current-Steering DACs

The high diversity of the DAC correction methods is partly due to the high diversity of the DAC errors, e.g. errors due to amplitude and timing mismatch, switching disturbances, transistor non-linearity, etc. Next to the nominal DAC signal transfer function (STF), these errors create a DAC error transfer function (ETF):

$$\underbrace{f\left(Inputs, i_1, i_2, i_3, \ldots, e_1, e_2, e_3, \ldots\right)}_{\text{D/A Function}} = \underbrace{f_S\left(Inputs, i_1, i_2, i_3, \ldots\right)}_{\text{Signal Transfer Function (STF)}} + \underbrace{f_E\left(Inputs, e_1, e_2, e_3, \ldots\right)}_{\text{Error Transfer Function (ETF)}}, \tag{5.1}$$

where *Inputs* represent the DAC inputs, e.g. input data, clock, control signals; i_j represents the nominal analog units, e.g. current cells and their switching; e_j represent the errors of the analog units. The DAC correction methods aim to minimize the contribution of the ETF, so that the DAC transfer function is as close as possible to the nominal STF. This section classifies DAC correction methods. It identifies missing methods and proposes new methods to fill in these gaps.

5.2.1 Classification

The paper classifies the DAC correction methods in "groups", "categories", and "classes", see Tables 5.1 through 5.3.

Three groups distinguish methods according to how they improve the DAC linearity. The first group *prevents* the errors (e_j) to have any influence on the DAC output and hence prevents the ETF (for those errors). This group is named Error Transfer Function Prevention (*ETFP*). Examples include Return-to-zero (RZ) and Differential-Quad Switching (DQS). The second group *corrects* the errors and hence corrects the ETF (Error Transfer Function Correction (*ETFC*)). Examples include self-calibration, errors mapping and DEM). The third group *compensates* the STF for the effect of the ETF (Signal Transfer Function Compensation (STFC)). Examples include digital pre-distortion and suppression of HD (introduced in this paper).

Three categories are defined along three different angles of incidence: *error measurement*, *redundancy*, and *system level*. Each category is further split up into two dichotomous classes, and the common characteristics of the methods of a specific class are derived.

Table 5.1 Error measurement category, with its two sub-classes

Group	With error measurements	Without error measurements
ETFP	N/A (the error is prevented from occurring; by definition, there can be no measurements)	Return-to-zero: [1, 2] Differential-quad switching: [3] Turned ON cascade switches: [4]
ETFC	Self-calibration: [5, 6], this work Mapping: [7, 8], this work Mapping for sub-binary radix DACs: [8]	DEM: [9] Segmented DEM (parallel sub-DAC level): [10]
STFC	Digital pre-distortion, [11, 12]	Suppression of HD, this work
	Common characteristics	**Common characteristics**
	Analog error measurement Background and foreground modes	No exact knowledge on the errors Background mode only

Table 5.2 Redundancy category, with its two sub classes

Group	Intrinsic redundancy	Extrinsic redundancy
ETFP	Differential-quad switching: [3]	Return-to-zero: [1, 2], Turned ON cascade switches: [4]
ETFC	Mapping: [7, 8], this work, DEM: [9, 10]	Self-calibration: [5, 6], this work
STFC	Suppression of HD: this work	Digital pre-distortion: [11, 12]
	Common characteristics	**Common characteristics**
	Risks of insufficient own resources for correction	Risks of deterioration of DAC intrinsic performance Complexity

Table 5.3 System-level category, with its two sub classes

Group	Low-level correction method	High-level correction method
ETFP	Return-to-zero current cell: [2] Differential-quad switching: [3] Turned ON cascade switches: [4]	Return-to-zero output stage, [1]
ETFC	Self-calibration: [5], this work Mapping for unary currents:. [7] Mapping for sub-binary radix DACs: [8] Input data reshuffling (DEM): [9]	Self-calibration: [6], Mapping for sub-DACs: this work Segmented DEM (parallel sub-DAC level): [10]
STFC	No examples for DACs Possibly, V-I converter linearization: [13]	Digital pre-distortion: [11, 12] Suppression of HD: this work
	Common characteristics	**Common characteristics**
	Increased hardware resources Complexity	Reduced dependence on DAC architecture

As a next step, similar techniques in the two classes of each category (for all groups) are identified. By doing so, the classification automatically points to methods that do not exist yet in the open literature. Based on the clues of the common properties and identified common techniques, this section then proposes exemplary solutions to fill in these gaps. All proposed new methods are based on parallel sub-DACs. Tables 5.1 through 5.3 show the proposed classification.

Table 5.1 shows the first category, looking from an error measurement view, and divides this category into two dichotomous classes: "*with error measurements*" and "*without error measurements*".

The mapping and DEM correction methods belong to the same ETFC group but to different classes. They are similar in the way they implement the correction but different in the way they use the error information. Both methods share the technique of rearranging the switching sequence of the DAC unary current cells. The mapping methods choose one static switching sequence, based on measured information for the errors e_j, while the DEM methods change the switching sequence randomly or periodically in time to average the errors e_j.

For the *DTFC* group, there is no counterpart known for the "digital pre-distortion" in the other class, so there is no corresponding method like mentioned earlier for the *ETFC* group. However, a new method to compensate the distortion components, mentioned as "Suppression of HD", is proposed and explained further in this paper. This method has been published independently, but then validated and applied in the RF field [14].

Table 5.2 shows the next category. It lists the DAC correction methods from a DAC redundancy point of view, and divides this category into two classes: *intrinsic* and *extrinsic* redundancy. Here, the intrinsic redundancy is defined as circuits which are inside the DAC core and directly used in its D/A function, but which possess hidden, unused potential. For example, the DAC unary currents represent intrinsic redundancy: they are directly used in the conversion, but still have the freedom to use different switching sequences. The extrinsic redundancy is defined here as circuits which are outside the DAC core and are only indirectly used in its D/A function, to help improve the performance. For example, calibration DACs, e.g. the CALDACs in Fig. 5.9, represent extrinsic redundancy: the D/A function can exist without them, but they can correct the errors of the DAC current sources. Like in the previous category, some shared similar techniques can be identified here too: DQS and RZ (reducing the data-dependence of the switching errors); Mapping/DEM and self-calibration (reducing amplitude and timing errors); suppression of HD and digital pre-distortion (reducing DAC HD components). "Suppression of HD", which is a new method, is explained further in this paper. For the known methods "Self-calibration" and "Mapping", this paper proposes new implementations that are independent from the DAC architecture.

Table 5.3 shows category three: it lists the DAC correction methods from a system-level point of view. It divides in either being *low*- or *high-level* methods. The *low-level* methods are applied to the DAC current cells, while the *high-level* methods are applied to the DAC as a whole. In the open literature, there are examples of RZ, self-calibration, and DEM methods implemented in both *low*- and *high-levels*.

However, no examples are found for the high-level mapping methods. Recalling that the DEM uses similar techniques as the mapping methods from Table 5.1, the potential high-level mapping method is supposed to be similar to the segmented DEM (based on parallel sub-DACs) but using error measurements instead of time averaging. This paper proposes such a high-level mapping method that is based on parallel sub-DACs.

5.2.2 New Correction Methods Based on Parallel Sub-DACs

As indicated already in Sect. 5.2.1, we propose using parallel current-steering sub-DACs to facilitate many DAC correction methods. An $N + log_2 M$ bit resolution DAC architecture based on M parallel N bit current-steering sub-DACs is shown in Fig. 5.1. The parallel sub-DACs can be considered as a special segmentation technique where multiple binary sets are allocated along with the MSB parts. Therefore, no excessive resources are required for the sub-DACs, compared to a normal $N + log_2 M$ bit resolution DAC. Recently, in an independent research [10], also two parallel sub-DACs to realize the segmented DEM method have been proposed.

An important advantage of all methods based on parallel sub-DACs is that the correction method is decoupled from the DAC architecture. Therefore, the architecture of the sub-DAC unit can be tailored for different targets. For example, the segmented DEM can use the area- and power-efficient binary architectures. The same argument is valid for the methods proposed in this paper: high-level mapping (Sect. 5.2.2.1), suppression of HD (Sect. 5.2.2.2), and binary currents calibration (Sect. 5.2.2.3).

5.2.2.1 New Method 1: High Level Mapping

Each of the M parallel N-bit sub-DACs can be considered as a separate DAC entity but when they are used together, system-level redundancy is added to the overall DAC system. The digital input word $w(nT)$ can be converted to an analog signal

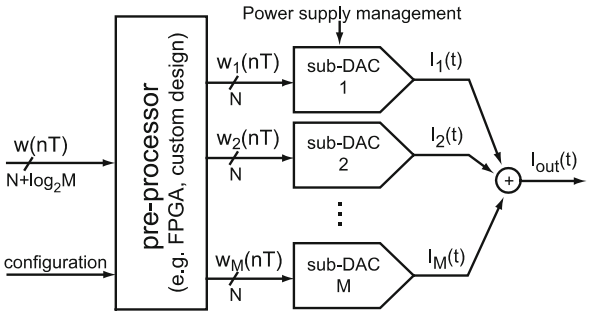


Fig. 5.1 A DAC architecture based on parallel sub-DAC units

i_{out} just as in a conventional DAC. However, there is an extra degree of freedom in choosing the sub-DAC input words $w_j(nT),\ j = 1, 2, \ldots, M$, thanks to the redundancy of the architecture. The digital pre-processor divides up the words $w(nT)$ among $w_j(nT)$. The sum of all $w_j(nT)$ should equal $w(nT)$:

$$w(nT) = w_1(nT) + w_2(nT) + \cdots + w_M(nT) \tag{5.2}$$

The output signal current is a sum of the sub-DAC output currents. The static and dynamic errors of these currents $i_{e_j}(t),\ j = 1, 2, \ldots, M$ also sum up and contribute to the output. Therefore, the output of the DAC can be considered as having an ideal part $i_{out}(t)$ and an error part $i_e(t)$:

$$i_{out}(t) + i_e(t) = i_{out1}(t) + i_{e1}(t) + \cdots + i_{outM}(t) + i_{eM}(t) \tag{5.3}$$

The mapping correction technique minimizes $i_e(t)$ by means of mutual error compensation. So far in the literature, the mapping technique is studied only for the unary MSB part of the segmented DAC, because the unary coding features intrinsic redundancy [15–17]. Since the binary coding features no intrinsic redundancy, the mapping techniques cannot be used at that level. However, the proposed DAC platform introduces redundancy at a higher level, the sub-DAC level, where sub-DACs are combined. Therefore, the high-level mapping can be used independently of the sub-DAC unit architecture.

Once the information of the particular current errors is available, the digital pre-processor can be programmed to distribute each individual code $w(nT)$ in such a way that all $i_{e_j}(t)$ for that code sum to minimum. Thus, the map selection criterion is:

$$\min[i_e(nT)] = \min[i_{e1}(nT) + i_{e2}(nT) + \cdots + i_{eM}(nT)] \tag{5.4}$$

Except for the initial and final codes of the DAC transfer characteristic, the number of possible combinations of $w_j(nT)$ to construct $w(nT)$ is very large.

5.2.2.2 New Method 2: Suppression of Harmonic Distortion

The DAC HD spurs for an input sinewave are a function of the DAC non-linear transfer characteristic, the input signal frequency f_{in} and the sampling frequency F_s. If F_s and f_{in} are known, the frequencies of the DAC HD components can be predicted:

$$f_{HD_k^p} = kf_{in} \pm pF_s, \tag{5.5}$$

where p is the p^{th} image band, and k is the k^{th} harmonic component. Consider now multiple sub-DACs working in parallel, as shown in Fig. 5.1. Let the pre-processor add particular phase delays $\varphi_{in}^{(m)}$, with $m \in [1, 2, \ldots, M]$, to the digital input signal of all sub-DACs. Then, for an input sinewave with frequency $\omega_{in} = 2\pi f_{in}$ and sampling frequency $\omega_s = 2\pi F_s$, these sub-DAC outputs are:

$$I_m(j\omega) = \sum_{p=-\infty}^{\infty} \sum_{k=1}^{\infty} A_{k,p}^{(m)} e^{j\left((k\omega_{in}+p\omega_s)t+p\varphi_s^{(m)}\right)} e^{jk\varphi_{in}^{(m)}}, \tag{5.6}$$

where $I_m(j\omega)$ is the output of the m^{th} sub-DAC in the complex domain, $A_{k,p}^{(m)}$ is the amplitude of the k^{th} harmonic component from the p^{th} image band of the m^{th} sub-DAC, $\varphi_s^{(m)}$ is the phase of the sampling signal of the m^{th} sub-DAC. When M sub-DACs operate in parallel, their combined output can be approximated as the superposition of their individual outputs. Let the M sub-DACs be synchronized, i.e. $\varphi_s^{(m)} = 0$ for all m, and let the M sub-DACs be identical, i.e. $A_{k,p}^{(m)} = A_{k,p}$ for all m. Then, the combined output $I_{out}(j\omega)$ of the M parallel sub-DACs is:

$$I_{out}(j\omega) = \sum_{m=1}^{M} I_m(j\omega) = \sum_{p=-\infty}^{\infty} \sum_{k=1}^{\infty} A_{k,p} e^{j(k\omega_{in}+p\omega_s)t} \sum_{m=1}^{M} e^{jk\varphi_{in}^{(m)}}, \tag{5.7}$$

A proper combination of $k\varphi_{in}^{(m)}$ can minimize the factor $\sum_{m=1}^{M} e^{jk\varphi_{in}^{(m)}}$ and hence suppress the combined effect, $A_{k,p}^{(1,2,\ldots,m)}$, for all p. That is to say that for a particular combination of the given sub-DAC non-linearity, i.e. the amplitudes $A_{k,p}^{(m)}$, and the input-signal phase shift, $\varphi_{in}^{(m)}$, the DAC output HD spurs can be suppressed through mutual compensation.

For example, suppression of HD_3 can be achieved with two nominally identical parallel DACs, i.e. $A_{k,p}^{(1)} = A_{k,p}^{(2)} = A_{k,p}$, converting phase shifted input signals by $\pi/3$, i.e. $\varphi_{in}^{(1)} = \varphi_{in}^{(2)} - \pi/3$, and combining their outputs to reduce the superposition of the 3rd harmonic amplitude $A_{3,p}^{(1,2)}$ for all image bands p:

$$A_{3,p}^{(1,2)} = A_{3,p}\left|e^{j(0\cdot\pi)} + e^{j\left(3\cdot\frac{\pi}{3}\right)}\right| = 0 \tag{5.8}$$

Figure 5.2 visualizes the example by means of phazor diagrams. In theory, $A_{1,p}^{(1,2)}$ (main signal) is increased, $A_{2,p}^{(1,2)}$ remains the same, and $A_{3,p}^{(1,2)}$ is cancelled. Thus,

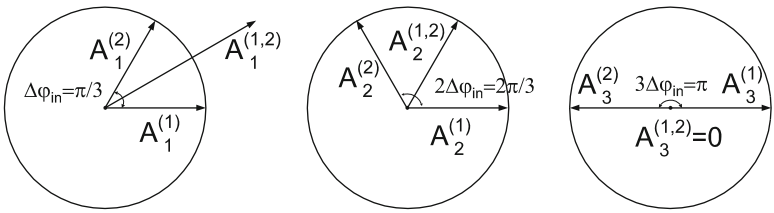


Fig. 5.2 Phazor diagrams for the main signal (A_1) and the second (A_2) and third (A_3) harmonic tones for two sub-DACs processing phase-shifted input signals

HD_2 is reduced and HD_3 is cancelled. In practice, due to the sub-DAC differences $A_{k,p}^{(1)} \neq A_{k,p}^{(2)}$ and error mechanisms located beyond the summation point, the reduction of both harmonic distortion spurs is less. Such error mechanisms include the mismatch errors and the finite output impedance of the current switching cells.

5.2.2.3 New Method 3: Self-Calibration of Binary Currents

In the literature so far, the calibration of currents is limited only to identical unit currents, e.g. unary currents [5]; or unary currents and a few binary currents built on units [2, 18]. The multiple parallel sub-DACs provide multiple binary sets of currents, which can be used with a calibration technique to calibrate all unary and binary currents.

The binary calibration algorithm is described with the calibration Eqs. 5.9 and 5.10. This algorithm can be used with different calibration methods, e.g. [5]. Three calibration loop steps are needed to correct the MSB binary currents $I(B)(1)$ and $I(B)(2)$, with B being the MSB binary current, the indexes (1) and (2) indicate sub-DAC 1 and 2. I_{ref_u} is the reference which is nominally twice larger than $I(B)$. In segmented DACs, I_{ref_u} is nominally equal to the unary current. Each equation includes an adjustable gray term and constant black terms. Steps 1 to 3 adjust the B bit binary currents of binary sets 1 and 2 to $0.5I_{ref\_u}$. After step 2, $I(B)(1)$ and $I(B)(2)$ are equal. Step 3 adjusts $I(B)(1)$ and $I(B)(2)$ at the same time, so they are made equal to $0.5I_{ref\_u}$.

$$\begin{aligned}
&1: I_{bin}(B)(1) + \sum_{i=1}^{B-1} I_{bin}(i)(1) + 1LSB \vdots\!\!\leftleftarrows I_{ref_u} \\
&2: I_{bin}(B)(2) + \sum_{i=1}^{B-1} I_{bin}(i)(1) + 1LSB \vdots\!\!\leftleftarrows I_{ref_u}, \\
&3: I_{bin}(B)(1) + I_{bin}(B)(2) \vdots\!\!\leftleftarrows I_{ref_u}
\end{aligned} \tag{5.9}$$

In like manner, the calibration flow continues with the $B-1$ currents $I(B-1)(1)$ and $I(B-1)(2)$. Their calibration is shown in Eq. 5.10. The left side of the equation for step 6 includes the already adjusted $I(B)(1) = 0.5I_{ref\_u}$. The sum $I(B-1)(1) + I(B-1)(2)$ is calibrated to $0.5I_{ref\_u}$, and hence $I(B-1)(1) = I(B-1)(2) = 0.25I_{ref\_u}$. The rest of the binary currents down to $I(1)(1)$ and $I(1)(2)$ are similarly calibrated.

$$\begin{aligned}
&4: I_{bib}(B-1)(1) + \sum_{i=1}^{B-2} I_{bib}(i)(1) + \sum_{i=B}^{B} I_{bib}(i)(1) + 1LSB \vdots\!\!\leftleftarrows I_{ref_u} \\
&5: I_{bib}(B-1)(2) + \sum_{i=1}^{B-2} I_{bib}(i)(1) + \sum_{i=B}^{B} I_{bib}(i)(1) + 1LSB \vdots\!\!\leftleftarrows I_{ref_u}, \\
&6: I_{bib}(B-1)(1) + I_{bib}(B-1)(2) + I_{bib}(B)(1) \vdots\!\!\leftleftarrows I_{ref_u}
\end{aligned} \tag{5.10}$$

5.3 Analysis of DAC Self-Calibration Methods

In this section, the self-calibration methods are further analyzed. This paper defines self-calibration as follows. "*Self*" means that the correction method is fully autonomous: the error measurement, the algorithm, and the error correction are integrated on-chip and no special activity from the customer is required. "*Calibration*" means a correction method that measures and corrects the DAC errors in the same domain (e.g. analog, time). This definition is meant to distinguish the self-calibration correction method from the other DAC correction methods for the sake of classifying their properties.

Self-calibration is a natural and powerful method to correct the DAC errors. This method directly targets the errors. It measures the errors, then processes the measurement information, and applies correction. Thus, the three defining blocks of self-calibration are self-measurement, integrated algorithm (error processing), and self-correction. Figure 5.3 shows a conceptual block diagram of a generalized self-calibrating current-steering DAC. The three self-calibration blocks are shown in boxes together with some of their main sub-elements.

5.3.1 Self-Measurement Block

The self-measurement block has three sub-elements: measurement infrastructure, measurement device, and the reference. The measurement infrastructure is the

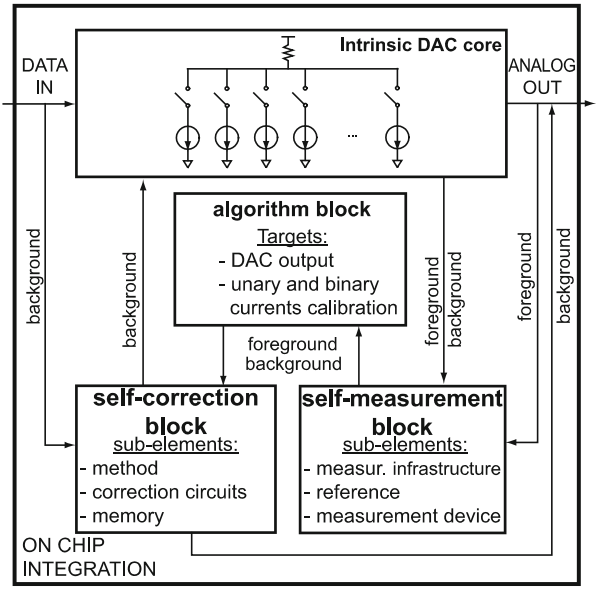


Fig. 5.3 Block diagram of DAC self-calibration

circuitry that makes possible the error measurement by the measurement device. The measurement infrastructure can be implemented in the upper [6], middle [2,5], or lower range [19,20] of the available voltage headroom. The measurement device quantifies the measured error with respect to the reference. The error information can be produced in either analog [2] or digital [5] domain. If only unary currents are calibrated, the mismatches between the reference and the non-calibrated binary currents must be minimized. There are two possibilities. The reference can be constructed with the non-calibrated binary currents, e.g. [5] or the binary currents as a group can be adjusted to the reference [2].

5.3.2 Algorithm Block

The algorithm block is responsible for implementing the sequence of operations that make possible the error measurement and its consequent correction. As shown in Fig. 5.3, the calibration algorithm can be executed in either background or foreground. The foreground algorithms are not active during the normal D/A conversion operations. The background algorithms form a loop that is being executed in parallel with the normal D/A conversion operations. To avoid frequency spurs at the output of the DAC, the background algorithms can be executed at randomized clock steps, as suggested in [21].

Figure 5.4 shows the self-calibration scheme and an 8-state FSM that controls it for *calibration of unary currents* with cancellation of the comparator input offset error. During ϕA, I_{temp} is calibrated to I_{ref}. X1 is the input calibration word for the CALDAC of I_{temp}. X1 is incremented until the comparator (formed by M2–M5) changes its output. The input offset error of the comparator is unavoidably recorded in I_{temp}, too. During ϕB, the unary currents for calibration $I_{u,k}$ are connected instead of I_{ref}. Each of them is calibrated to I_{temp}. This is done by incrementing the input calibration word of the CALDACs of $I_{u,k}$ X2 until the comparator changes its output. Since the input offset error is inversed, all $I_{u,k}$ are calibrated to I_{ref} free of the input offset error. To reduce the post-calibration errors, the polarity of the calibration quantization error is controlled in both ϕA and ϕB.

To calibrate *binary currents*, [2,18] suggest to introduce a calibrated unit element (CUE) current. Then, the unary and some of the binary currents are constructed with CUE. The calibration is executed e.g. as shown in Fig. 5.4. Those binary currents that are constructed with CUEs are effectively calibrated, too. Figure 5.5 shows the three different ways to choose the CUE with respect to the DAC segmentation. For high DAC accuracy, the calibration of some of the binary currents is possible with option (a). Option (b) provides design compactness, since the unary currents coincide with the CUEs. Option (c) requires less area for the CALDACs, since the CUE is defined over multiple unary current cells and hence the CALDACs are shared.

Here, a new algorithm for *full calibration of binary currents* is proposed. The main principle is demonstrated with Eqs. 5.9 and 5.10. Figure 5.6 shows a

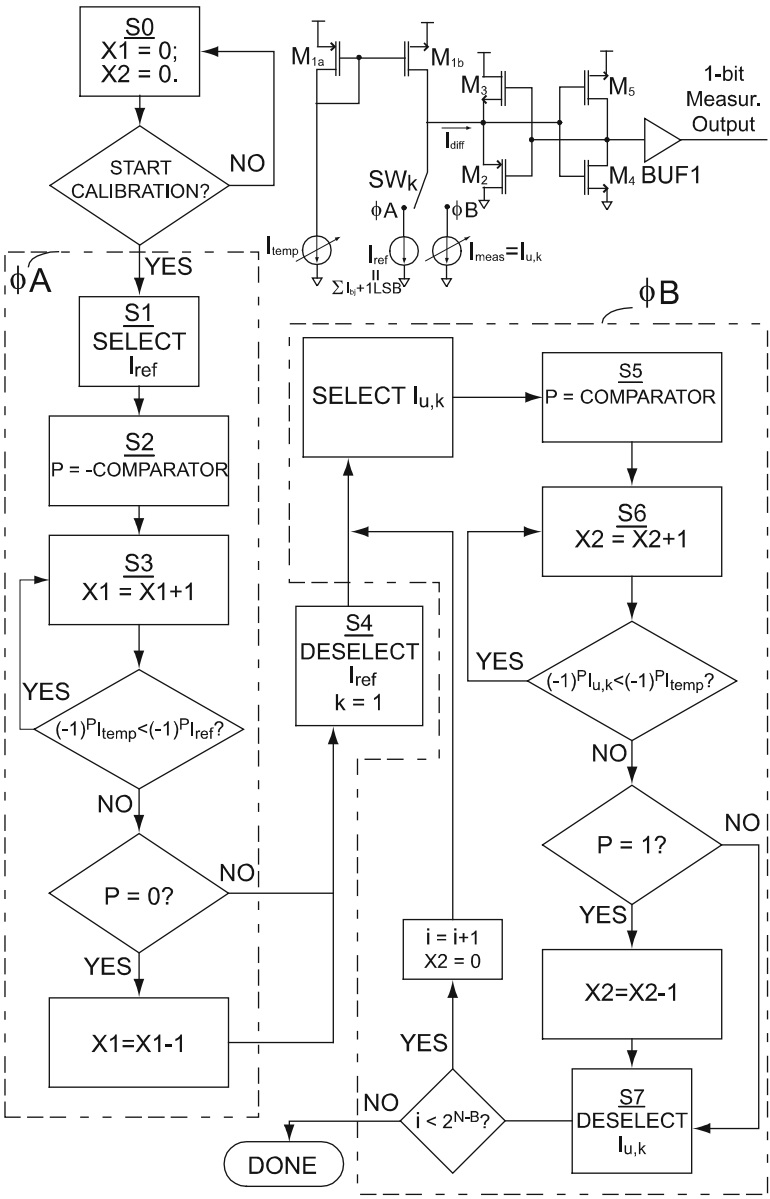


Fig. 5.4 FSM chart of the unary currents self-calibration algorithm, based on [5]

self-calibration scheme and the FSM that controls it for calibration of binary currents. The shown FSM is for two sets of binary currents m = 1 and m = 2, e.g. from two sub-DACs. The FSM can be easily adapted to more binary sets. Since all currents are calibrated no special requirements are set to the construction of the

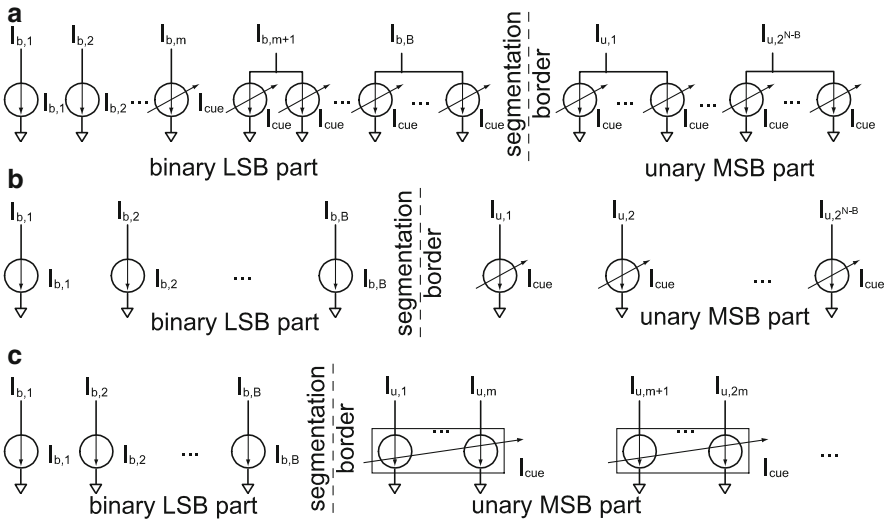


Fig. 5.5 Three different options for the CUE bit in a segmented DAC, from [18]: (**a**) in the LSB binary part (**b**) at the unary level (**c**)in the MSB unary part

reference current and to the compensation of the input offset errors of the comparator (formed by M2–M5). The FSM implements the calibration equations in an iterative way. Each cycle is formed by three steps. In the first two steps, the binary currents from the two sets are made equal to each other. In the third step m = 3, both simultaneously are made equal to an exact portion of the reference.

5.3.3 Self-Correction Block

The calibration self-correction block concerns the means for compensating the measured errors. It has three sub-elements: the sub-method, the correction circuit, and the correction memory.

The self-correction sub-method realizes the correction of the measured error. The correction can be applied either at high (DAC system) [6] or low (DAC elements) [5]. Further, the correction current can be injected [5] or the main quantity can be regulated [23]. Finally, the correction quantity can be either discrete [5] or continuous [2]. The different options are illustrated in Fig. 5.7.

The two main choices for the correction circuits are the calibrating DACs (CALDACs) attached to every current source and the circuits regulating the V_{gs} of the current sources. In case of a CALDAC, the calibration memory is digital. In the case of V_{gs} regulation the calibration memory can be either digital or analog (capacitor).

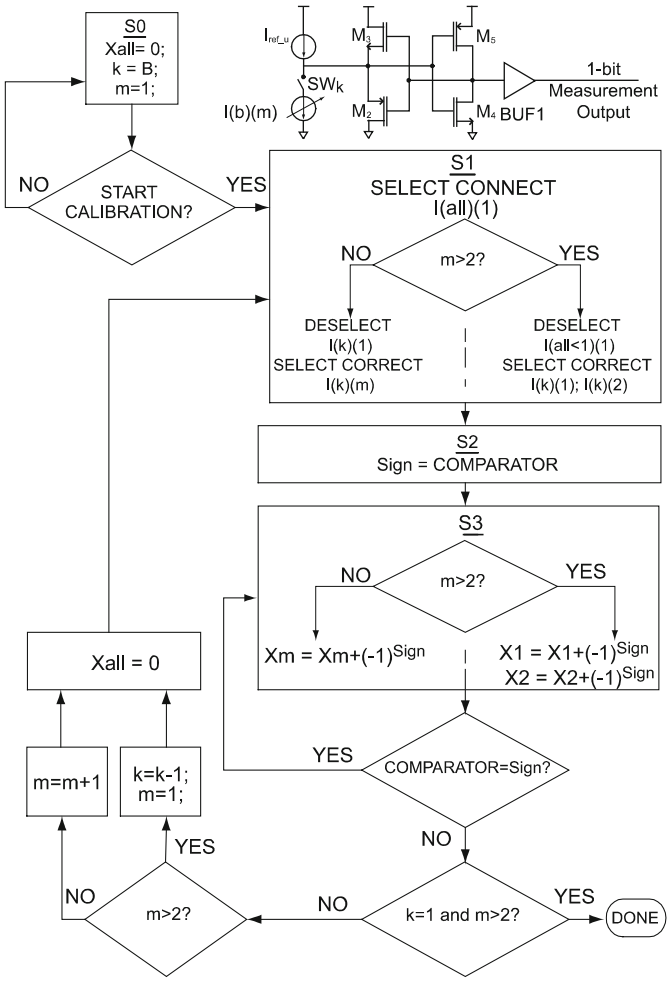


Fig. 5.6 FSM chart of the binary currents self-calibration algorithm, based on [22]

5.4 Parallel Current-Steering DACs for Flexibility and Smartness

Flexibility is a concept introduced to adapt the electronic systems to their ever changing environment which includes the new market realities, the application requirements and the IC technologies. Flexibility is already a very successful concept in the area of digital electronics. The programmable FPGA market is continually expanding and naturally it should cross into the mixed-signal domain. The SoC (System-On-Chip) co-integration between FPGAs and mixed-signal interfaces will introduce the programmability advantages into the mixed-signal domain, e.g. decoupling the application from the silicon design, reducing time-to-market, and bringing application tailored performance at minimum cost.

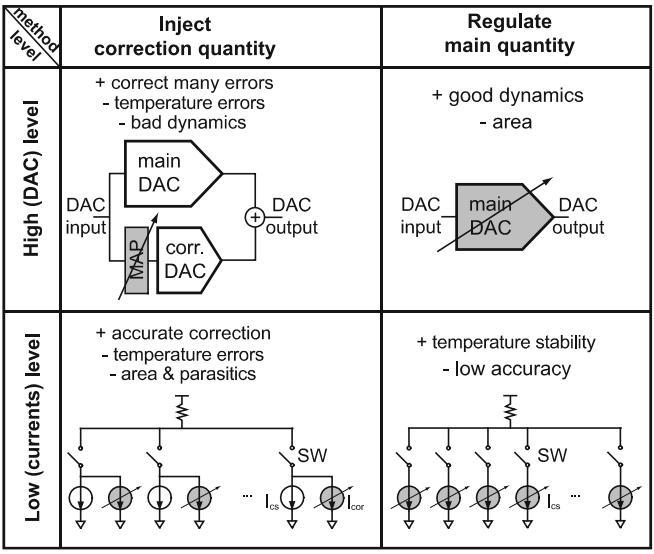


Fig. 5.7 Classification of the DAC self-calibration methods with respect to either injecting correction or regulating the main quantity

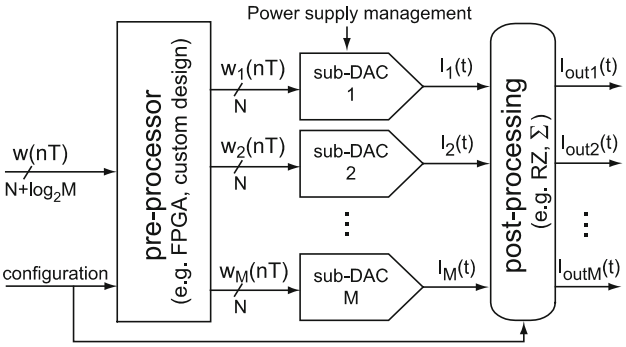


Fig. 5.8 A flexible DAC architecture based on parallel sub-DAC units

A new flexible architecture for current-steering DACs is proposed here. Its main principle suggests introducing DAC redundancy at system level. In practice, this principle is translated into using parallel low-resolution sub-DAC units as building blocks of a high resolution flexible DAC platform. A digital pre-processor distributes the digital input among the sub-DACs, in a way controlled by the operation mode (op-mode) of the DAC platform. The sub-DAC outputs are combined in the analog post-processing block, which in its simplest form is just a summation, like shown in Fig. 5.1 where the self-correction properties of this architecture are discussed. Here, the focus is on the flexibility properties of the parallel sub-DACs. Figure 5.8 shows a diagram of the flexible DAC concept.

The basic building block is an N-bit sub-DAC. If M sub-DACs are used in parallel, then the overall resolution can range from N to $N + \log_2 M$ bits. The main features of this DAC platform include flexible design, flexible functionality, and flexible performance. At design level, the engineers have the flexibility of a modular design approach. At functional level, the customer can configure the platform to operate as either a single or multiple independent DACs. At performance level, the customer can program the DAC resolution, and power consumption, and do trade-offs between the DAC speed, linearity, and resolution. In addition, this architecture facilitates a number of DAC correction methods based on parallel sub-DACs, such as complete current sources calibration, error mapping, DEM and harmonic distortion cancellation. Exemplary operation modes (op-modes) are discussed in Sect. 5.5.2, where some design and measurement results are presented for four 12-bit parallel sub-DACs.

5.5 Design Examples and Measurements

This section presents two test-chip implementations that demonstrate in practice the concepts of self-calibration and flexibility.

5.5.1 Unary Currents Self-Calibration in a 12-bit 250 nm DAC

A 12-bit self-calibrated current-steering DAC is implemented in a standard 250 nm 1P5M CMOS process with a power supply of 2.5 V. The self-measurement infrastructure is based on the principle of current deviation at the second cascade as shown in Fig. 5.9 (a high-level block diagram of the self-calibrated DAC). To reduce the occupied silicon area, the DAC core was designed for 10-bit accuracy. It has 6–6

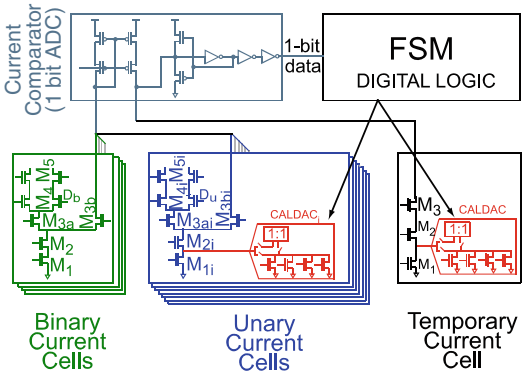


Fig. 5.9 Self-calibration DAC scheme for unary currents calibration, [5]

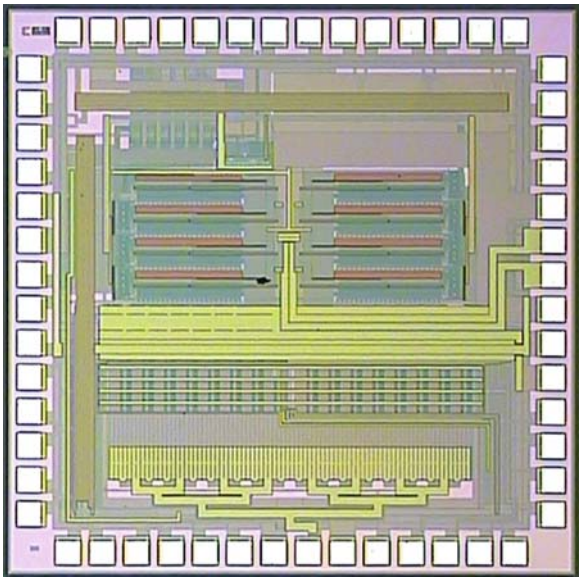


Fig. 5.10 A micrograph of the 250 nm self-calibrated DAC

segmentation and only the 63 MSB unary current sources are calibrated. The 6 LSB binary currents are not calibrated but they are used to construct the reference current.

The realized 12-bit self-calibrated DAC is shown in Fig. 5.10. The dimensions of the DAC core are 1.16 by 0.98 mm. The presented measurement results are for 20 mA full-scale current terminated on a 50 Ω differential load resistance. Figure 5.11 shows the measured INL and DNL of the DAC before and after self-calibration.

To isolate only the calibration improvement from the overall DAC performance and hence to estimate the potential of the calibration method, 1,827 unary current sources from different chip samples are measured before and after calibration. Figure 5.12 shows the measurement results as a distribution diagram of the relative LSB accuracy, normalized at the 12-bit level. The calibration of the unary currents exceeds the 14-bit level.

The calibration improves the dynamic linearity of the DAC, too. Figure 5.13 shows the DAC output frequency spectrum for input tone signal $f_{in} = 5\,\text{MHz}$, sampled at $F_s = 50\text{MS/s}$ before calibration. The SFDR is limited to about 68dB. Figure 5.14 shows the DAC output frequency spectrum after the calibration. The current mismatch errors are corrected and the DAC performance is improved to SFDR = 81dB. Note that the main harmonic spurs are well suppressed. SFDR is limited by high frequency spurs that are mainly related to both the non-calibrated binary currents and the dynamic characteristics of the particular DAC implementation. To isolate only the unary currents calibration improvement from the overall DAC

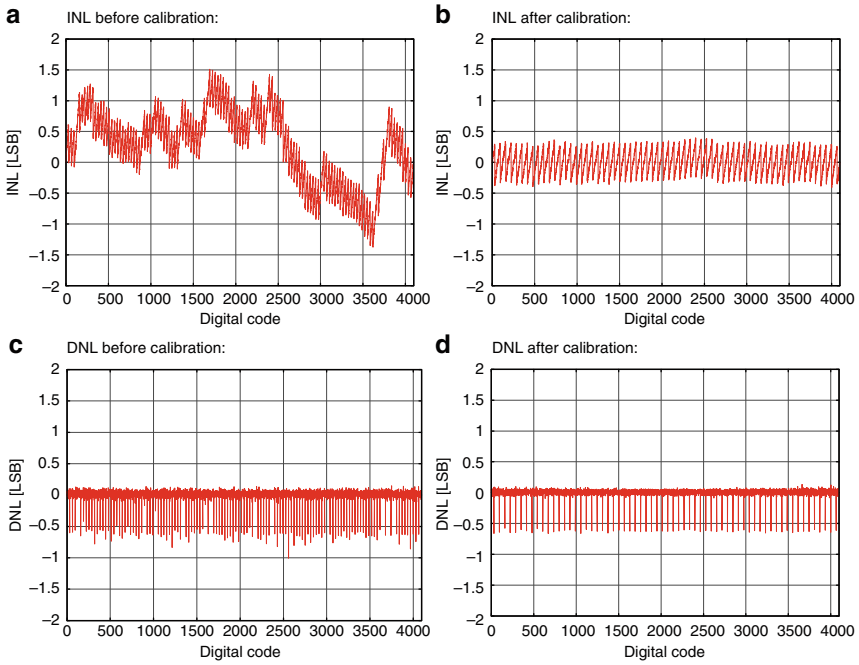


Fig. 5.11 Measured DAC INL and DNL. **(a)** INL before calibration, (**b**) INL after calibration, **(c)** DNL before calibration, **(d)** DNL after calibration

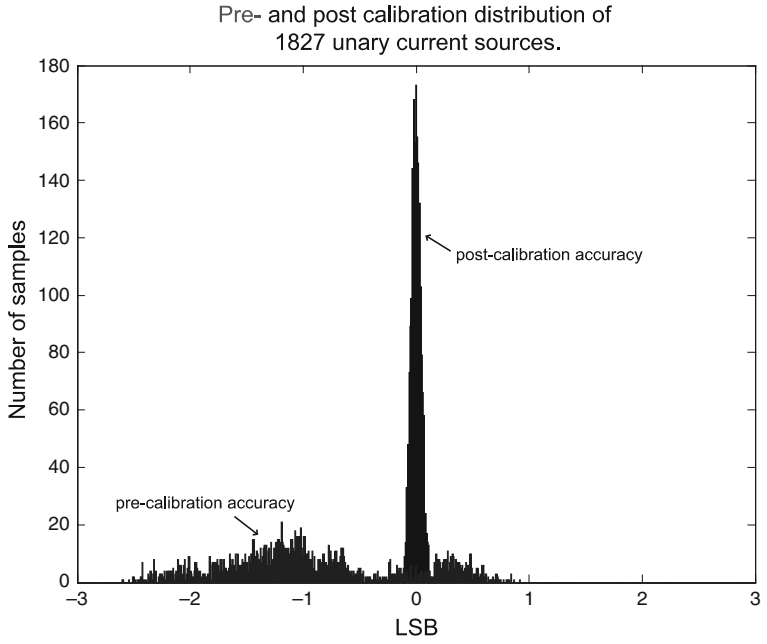


Fig. 5.12 Distribution of the accuracy (at 12-bit level) of 1827 measured unary current sources, before and after calibration

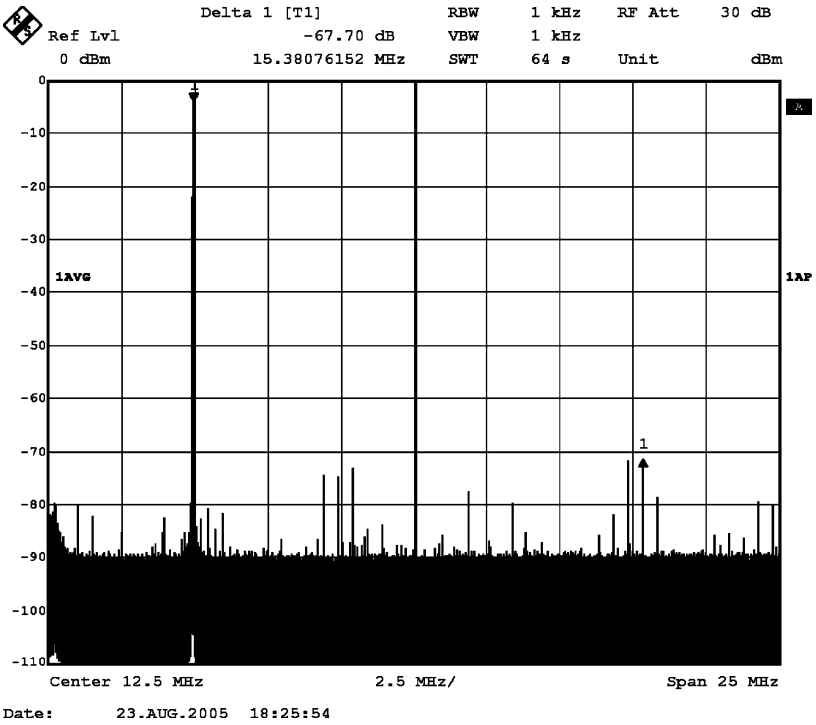


Fig. 5.13 DAC output spectrum, before calibration, SFDR = 68 dB

dynamics and hence to estimate the potential of the calibration method, the HD2, HD3, HD4, and HD5 (mainly due to the unary currents mismatch) are evaluated against f_{in}.

Figure 5.15 shows HD2 through HD5, before and after calibration. For lower $f_{in} < 3$ MHz, the calibrated HD components are at about level −85 dB, while the worst HD before calibration is at about level −66 dB. Thus, the calibration improvement is almost 20 dB, i.e. more than 3 bits, which is also suggested by the static distribution of the calibrated unary currents in Fig. 5.12. Beyond $f_{in} > 7$ MHz, the dynamic errors dominate the DAC performance and the advantage of the static errors calibration are reduced.

While at lower frequencies $f_{in} < 6$ MHz, the DAC SFDR before calibration is limited by the first few harmonic distortion components (see Fig. 5.13), the DAC SFDR after calibration is limited by some higher frequency spurs. Along with the particular DAC realization, the non-calibrated binary currents contribute to the power of these high frequency components. A close look at the DNL and INL characteristics (see Fig. 5.11) reveals that the non-calibrated errors of the binary currents would produce small high-frequency modulation products. Figure 5.16 shows the DAC SFDR against the input frequency tones. The DAC SFDR shows about 80 dB linearity for frequencies up to 5 MHz.

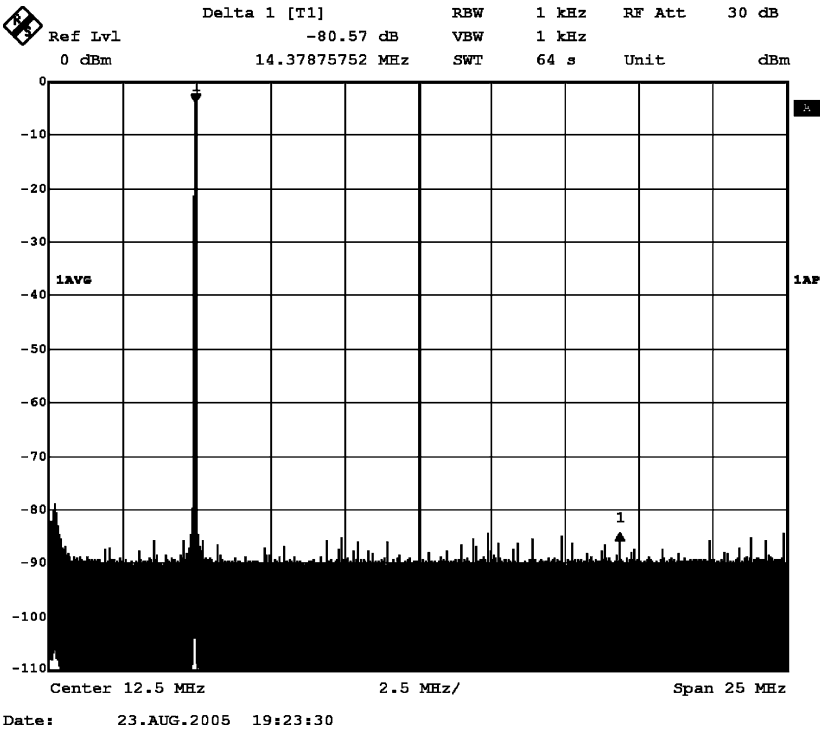


Fig. 5.14 DAC output spectrum after calibration, SFDR = 81 dB

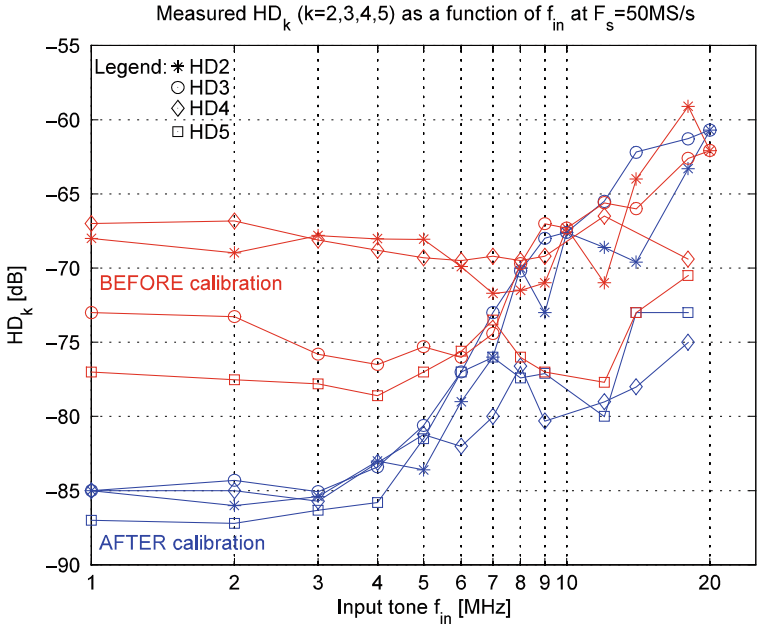


Fig. 5.15 DAC Harmonic distortion (HD) components against input frequency, before and after calibration

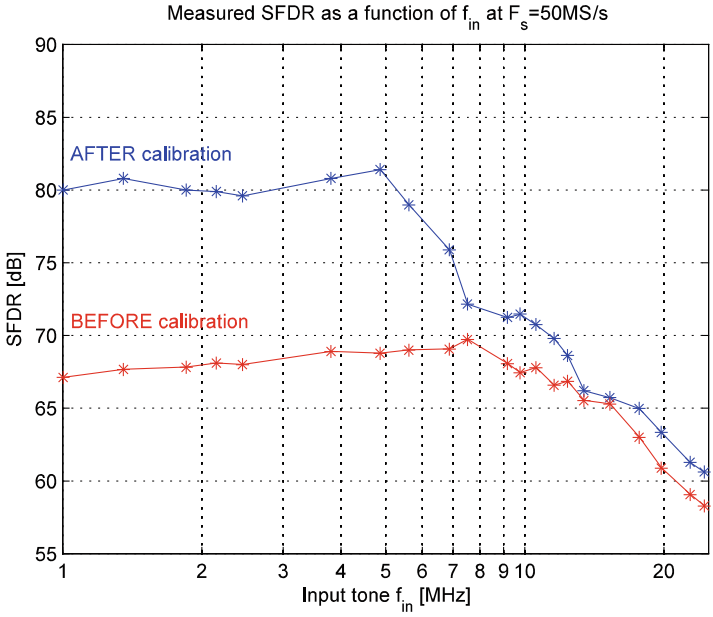


Fig. 5.16 DAC SFDR against the input tone frequency

5.5.2 Both Unary and Binary Currents Self-Calibration in a 12-bit 180 nm Quad-Core Flexible DAC

The concepts of the binary currents calibration are implemented in a quad-core 12-bit DAC in an industry standard 180 nm 1P6M CMOS process. The DAC is based on four parallel 12-bit sub-DAC units implemented with a segmented architecture of 8-bit binary LSB and 4-bit unary MSB (15 unary MSB currents). The particular advantages of this flexible architecture are presented in [24]. The implementation includes full on-chip-integrated all-currents calibration, based on the algorithms shown in Figs. 5.4 and 5.6. The binary currents calibration method uses the four sets of binary currents of the four sub-DAC cores to implement the binary calibration algorithm. A micrograph of the test chip is shown in Fig. 5.17.

The four sub-DACs A, B, C, and D are implemented horizontally next to each other. Part 1 indicates the array of coarse current sources M1. Part 2 indicates the array of CALDACs. Part 3 indicates the array of cascade transistors, the calibration switches, and the digital calibration logic. Part 4 indicates the synchronization latches implemented with Current-Mode-Logic (CML) like in [15]. Part 4 includes also the CML binary-to-unary decoder for the sub-DAC unary MSB part. Part 5 is the CML pre-processor that allocates the input digital words to the sub-DACs, acting as a demultiplexor of the chip input. Part 6 is the input data LVDS buffers block and decoupling capacitors. Finally, part 7 is the reference current. The overall DAC area is only 0.8 mm^2, hence 0.2 mm^2 per 12 b sub-DAC unit. Comparatively

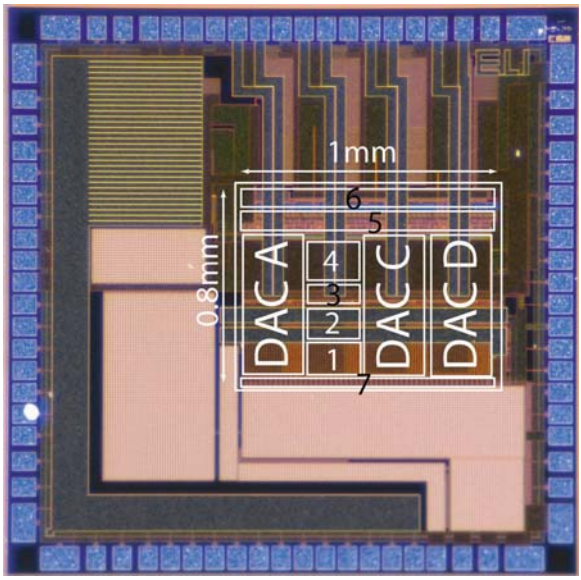


Fig. 5.17 Micrograph of the flexible self-calibrated quad-core DAC

to the literature, it is one of the smallest published designs. The area of the DAC signal current sources is small, because they are not designed for extreme matching. Their design is relaxed. It takes into account that the DAC calibration can improve their accuracy to more than 14-bit. In addition, the DAC implements segmentation with a large portion of binary LSB bits: 8. This reduces the overall area, while the non-linearity drawbacks of the binary bits are answered with the calibration.

The chip power consumption consists of three parts: start-up (calibration), digital (data processing), and analog (output signal) power consumption. The power for calibration is practically zero, because the calibration is run once at chip start-up and the results are memorized; during normal DAC operation the CMOS calibration logic is not active. Regarding the DAC analog signal output power, the presented results are measured for 24 mA full-scale current, i.e. 6 mA per sub-DAC, terminated with a 50 Ω differential load resistance. The full digital data processing power consumption of the flexible DAC is 118.8 mW at 1.8 V supply (66 mA current). The distribution of the digital power consumption is 27 mW per sub-DAC and 10.8 mW for the pre-processor. Thus, the overall DAC power consumption is flexible from 37.8 mW (only one sub-DAC is used and the rest are turned off) to 118.8 mW (all sub-DACs operate simultaneously).

An exemplary characterization of the DAC static performance is shown in the INL and DNL plots of Fig. 5.18. The calibration improves the DAC accuracy to an almost 14-bit level.

Figure 5.19 shows the measured DNL_{max} and INL_{max} for different amplitudes of the LSB correction step I_{cor}, given relatively to the LSB current of the DAC. This is the size of the LSB correction step of the unary CALDAC. The correction

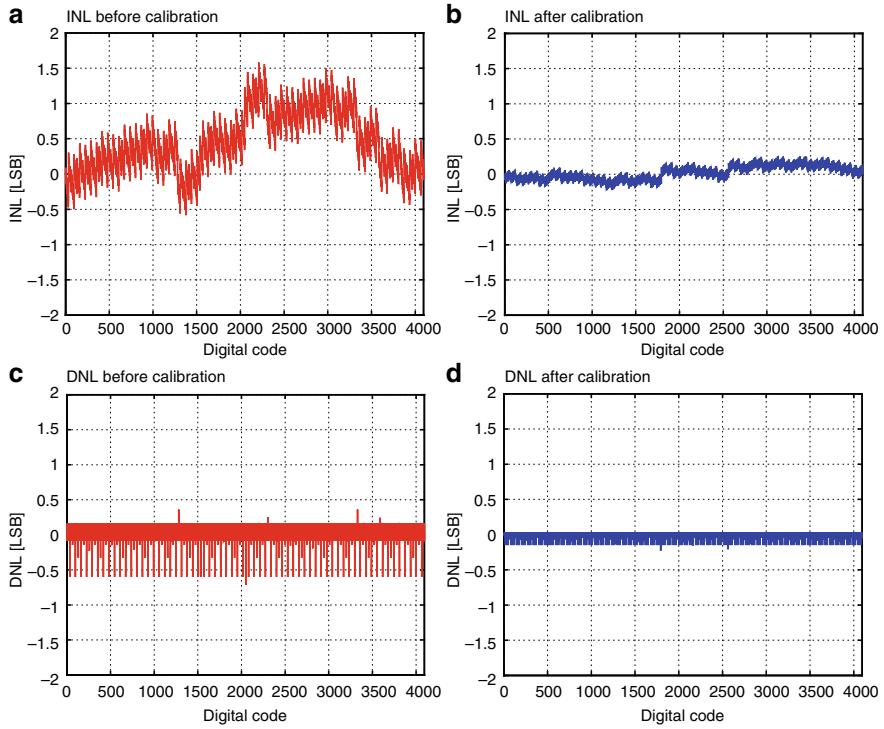


Fig. 5.18 Measured INL and DNL of four parallel sub-DACs (**a**) INL before calibration, (**b**) INL after calibration, (**c**) DNL before calibration, (**d**) DNL after calibration

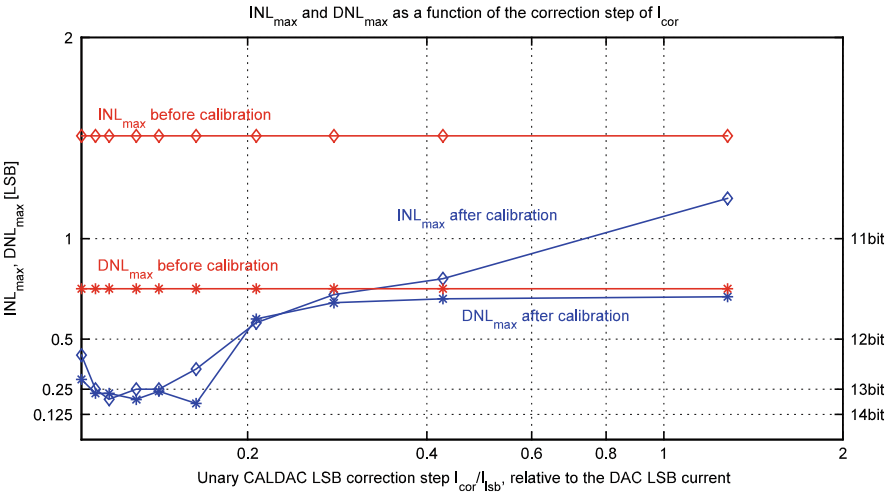


Fig. 5.19 Measured INL_{max} and DNL_{max} as a function of the LSB correction step of the unary CALDAC

steps of the binary CALDAC scale down by a factor of $\sqrt{2}$. Both DNL_{max} and INL_{max} depend on the size of the correction step I_{cor}. *The DAC accuracy depends on a design parameter I_{cor} and not on the tolerances of the IC fabrication process!* Small I_{cor} guarantees small correction error and hence high post-calibration DAC accuracy. However, if I_{cor} is too small the full-scale range of the CALDAC cannot cover the errors of both the calibrated DAC currents and the measurement device. Thus, some currents cannot be completely calibrated and their post-correction errors deteriorate the DAC accuracy, as shown in Fig. 5.19 for I_{cor} approaching small values.

Figure 5.20 shows the DAC SFDR performance for different op-modes, with and without self-calibration. At lower input frequencies f_{in}, more than 10 dB better intrinsic SFDR performance is achieved when all four sub-DACs operate together in parallel. At high f_{in}, more than 10 dB better intrinsic performance is achieved when only one sub-DAC is used. The main reason for this flexible performance is that different error mechanisms dominate at low and high speeds. For low f_{in}, the predominant error mechanism is the current mismatch error, i.e. the DAC amplitude accuracy. As more sub-DACs operate in parallel the current mismatch errors (or the post-calibration errors), both random and systematic, average out. For high f_{in}, the predominant error mechanisms arise due to synchronization errors of the current cells and the dynamic disturbances on the power lines and transistor biasing nets. Therefore, the use of fewer elements is advantageous, i.e. only one sub-DAC op-mode. The calibration of the DAC current sources improves further the DAC SFDR performance to above 80 dB. For frequencies up to 11 MHz, the predominant problem is the DAC current mismatches. Beyond 11 MHz the dynamic error mechanisms

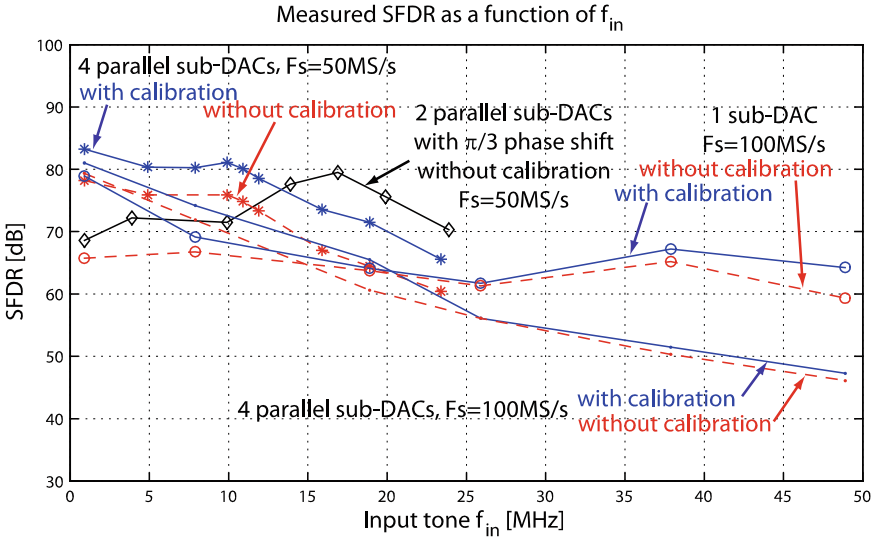


Fig. 5.20 SFDR performance against for different op-modes

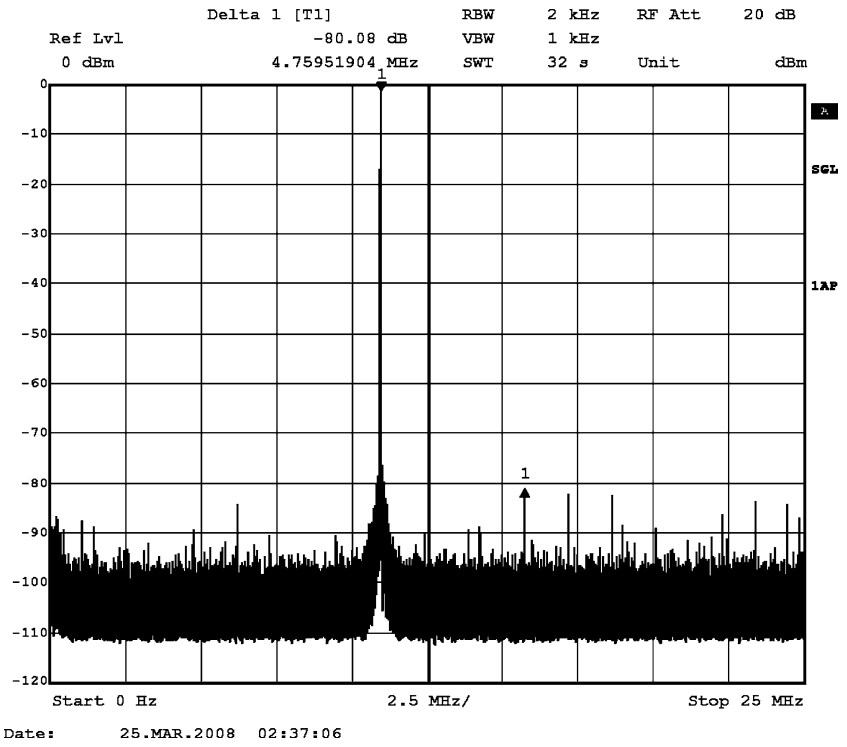


Fig. 5.21 DAC output spectrum for $f_{in} = 11\,\text{MHz}$ after calibration, SFDR = 80 dB

overshadow the calibration advantages. Figure 5.21 shows the DAC output spectrum after calibration for $f_{in} = 11\,\text{MHz}$.

The DAC performance for the suppression of HD method is shown with black line in Fig. 5.20, where two parallel sub-DACs are used. They convert $\pi/3$ phase-shifted replicas of f_{in}. Figure 5.22 shows the DAC output spectrum for $f_{in} = 16.9\,\text{MHz}$ and $F_s = 50\text{MS/s}$. All multiples of HD2 and HD3 are greatly suppressed. The dominant HD spurs are the multiples of HD5 and HD7, which is not suppressed by $\pi/3$ phase-shifting. HD7 limits the DAC SFDR.

5.6 Conclusions

A classification of the DAC correction methods with emphasis on the self-calibration is introduced. It structures the available knowledge from the open literature and indicates where new DAC correction methods can be developed. Three new DAC correction methods based on parallel current-steering sub-DACs are described: high-level mapping, suppression of DAC HD, and self-calibration of binary currents. The parallel sub-DACs architecture also realizes flexibility in

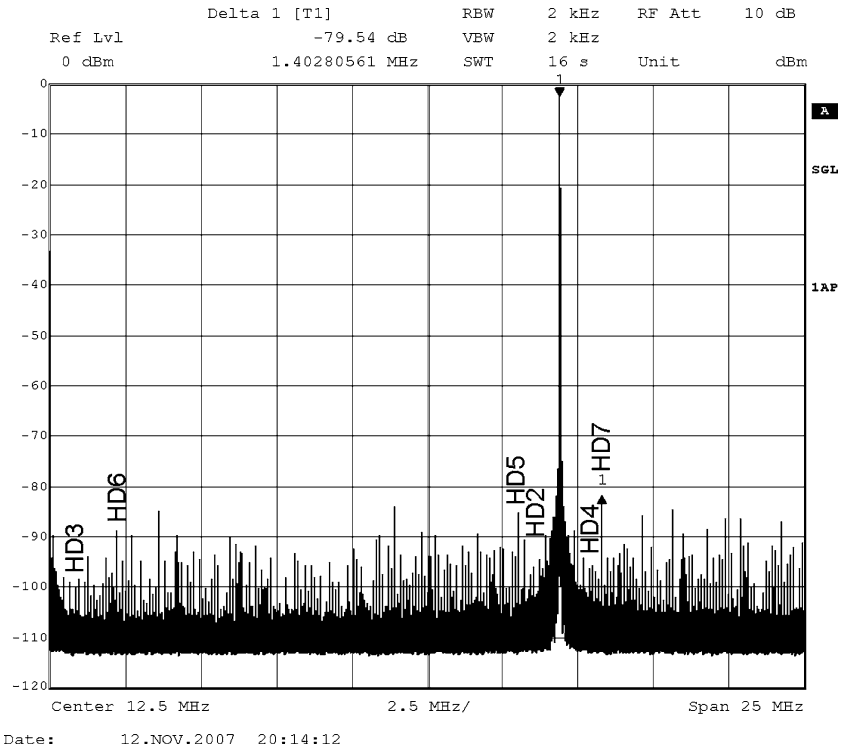


Fig. 5.22 DAC output spectrum for two sub-DACs in parallel processing a phase shifted by $\pi/3$ $f_{in} = 16.9\,\text{MHz}$, $F_s = 50\text{MS/s}$

DAC design, functionality, and performance. Two test-chip implementations validate the presented concepts. A 12-bit 250 nm self-calibrated DAC demonstrates a more than 14-bit accuracy of the calibrated unary currents and SFDR $>$ 80dB for $f_{in} = 5\,\text{MHz}$. The test-chip measurements indicate that to further improve the DAC performance and to decouple the DAC architecture from the self-calibration method, the binary currents should be calibrated, too. A 12-bit quad-core 180 nm self-calibrated flexible DAC test-chip demonstrates complete self-calibration for both the unary and binary currents, achieving SFDR $>$ 80 dB for $f_{in} = 11\,\text{MHz}$. In addition, the test chip demonstrates flexibility in different op-modes. A special op-mode implements the new method for suppression of DAC HD, achieving SFDR $= 79\,\text{dB}$ for $f_{in} = 16.9\,\text{MHz}$. With the robustness of the self-calibration and the flexibility of the proposed DAC architecture, the presented concepts are expected to assist the FPGA systems in their inevitable expansion in the mixed-signal domain.

References

1. A. R. Bugeja, B. S. Song, P. L. Rakers, and S. F. Gillig, "A 14-b, 100-MS/s CMOS DAC designed for spectral performance," *IEEE Journal of Solid-State Circuits*, vol. 34, pp. 1719–1732, 1999.
2. M. Clara, W. Klatzer, D. Gruber, A. Marak, B. Seger, and W. Pribyl, "A 1.5 V 13bit 130–300MS/s self-calibrated DAC with active output stage and 50 MHz signal bandwidth in 0.13μm CMOS," in *ESSCIRC 2008 Solid-State Circuits Conference, 2008. 34th European*, pp. 262–265, 2008.
3. P. Sungkyung, K. Gyudong, P. Sin-Chong, and K. Wonchan, "A digital-to-analog converter based on differential-quad switching," *IEEE Journal of Solid-State Circuits*, vol. 37, pp. 1335–1338, 2002.
4. C.-H. Lin, F. van der Goes, J. Westra, J. Mulder, Y. Lin, E. Arslan, E. Ayranci, X. Liu, and K. Bult, "A 12b 2.9GS/s DAC with IM3 $<$ $-$60dBc Beyond 1 GHz in 65 nm CMOS," in *Solid-State Circuits Conference, 2009. Digest of Technical Papers. ISSCC. 2009 IEEE International*, 2009.
5. G. I. Radulov, P. J. Quinn, H. Hegt, and A. van Roermund, "An on-chip self-calibration method for current mismatch in D/A converters," in *Solid-State Circuits Conference, 2005. ESSCIRC 2005. Proceedings of the 31st European*, pp. 169–172, 2005.
6. Y. Cong and R. L. Geiger, "A 1.5-V 14-bit 100-MS/s self-calibrated DAC," *IEEE Journal of Solid-State Circuits*, vol. 38, pp. 2051–2060, 2003.
7. K. Doris, "High-speed D/A converters: from analysis and synthesis concepts to IC implementation," in *Eindhoven University of Technology, Faculty of Electrical Engineering*. vol. Ph.D. thesis degree Eindhoven: Eindhoven University of Technology, 2004.
8. P. Harpe, J. M. Meulmeester, A. J. Hegt, and A. van Roermund, "Novel digital pre-correction method for mismatch in DACs with built-in-self measurement," *Proceedings of IEEE ADDA 2005*, 2005.
9. L. R. Carley and J. Kenney, "A 16-bit $4'\tau\eta$ order noise-shaping D/A converter," in *Custom Integrated Circuits Conference, 1988, Proceedings of the IEEE 1988*, pp. 21.7/1–21.7/4, 1988.
10. K. L. Chan, N. Rakuljic, and I. Galton, "Segmented dynamic element matching for high-resolution digital-to-analog conversion," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 55, pp. 3383–3392, 2008.
11. G. R. Spalding and R. L. Geiger, "Digital correction for improved spectral response in signal generation systems," in *1993 IEEE International Symposium on Circuits and Systems, ISCAS '93*, pp. 132–135, 1993.
12. E. Lopelli, J. D. van der Tang, and A. H. M. van Roermund, "A 1 mA ultra-low-power FHSS TX front-end utilizing direct modulation with digital pre-distortion," *IEEE Journal of Solid-State Circuits*, vol. 42, pp. 2212–2223, 2007.
13. S. Ouzounov, E. Roza, J. A. Hegt, G. van der Weide, and A. H. M. van Roermund, "A CMOS V-I converter with 75-dB SFDR and 360-μW power consumption," *IEEE Journal of Solid-State Circuits*, vol. 40, pp. 1527–1532, 2005.
14. E. Mensink, E. A. M. Klumperink, and B. Nauta, "Distortion cancellation by polyphase multipath circuits," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 52, pp. 1785–1794, 2005.
15. K. Doris, J. Briaire, D. Leenaerts, M. Vertregt, and A. van Roermund, "A 12b 500MS/s DAC with $>$70 dB SFDR up to 120 MHz in 0.18um CMOS," in *Solid-State Circuits Conference, 2005. Digest of Technical Papers. ISSCC. 2005 IEEE International*, 2005.
16. T. Chen, P. Geens, G. Van der Plas, W. Dehaene, and G. Gielen, "A 14-bit 130-MHz CMOS current-steering DAC with adjustable INL," in *Solid-State Circuits Conference, 2004. ESSCIRC 2004. Proceeding of the 30th European*, 2004.
17. T. Yongjian, H. Hegt, A. van Roermund, K. Doris, and J. Briaire, "Statistical analysis of mapping technique for timing error correction in current-steering DACs," in *IEEE International Symposium on Circuits and Systems, 2007. ISCAS 2007*, pp. 1225–1228, 2007.

18. G. I. Radulov, P. J. Quinn, J. A. Hegt, and A. H. M. van Roermund, "A start-up calibration method for generic current-steering D/A converters with optimal area solution," in *ISCAS 2005. IEEE International Symposium on Circuits and Systems, 2005*, vol. 1, pp. 788–791, 2005.
19. A. R. Bugeja and S. Bang-Sup, "A self-trimming 14-b 100-MS/s CMOS DAC," *IEEE Journal of Solid-State Circuits*, vol. 35, pp. 1841–1852, 2000.
20. H. Qiuting, P. A. Francese, C. Martelli, and J. Nielsen, "A 200MS/s 14b 97 mW DAC in 0.18/spl mu/m CMOS," in *Solid-State Circuits Conference, 2004. Digest of Technical Papers. ISSCC. 2004 IEEE International*, pp. 364–532, 2004.
21. M. Clara, W. Klatzer, B. Seger, A. Di Giandomenico, and L. Gori, "A 1.5 V 200MS/s 13b 25 mW DAC with randomized nested background calibration in 0.13/spl mu/m CMOS," in *Solid-State Circuits Conference, 2007. ISSCC 2007. Digest of Technical Papers. IEEE International*, pp. 250–600, 2007.
22. G. Radulov, P. Quinn, A. J. Hegt, and A. van Roermund, "Method and apparatus for calibrating a scaled current electronic circuit," in *US Patent 7,466,252*: Xilinx, 2008.
23. M. P. Tiilikainen, "A 14-bit 1.8-V 20-mW 1-mm^2 CMOS DAC," *IEEE Journal of Solid-State Circuits*, vol. 36, pp. 1144–1147, 2001.
24. G. I. Radulov, P. J. Quinn, P. Harpe, H. Hegt, and A. van Roermund, "Parallel current-steering D/A Converters for Flexibility and Smartness," in *Circuits and Systems, 2007. ISCAS 2007. IEEE International Symposium on*, pp. 1465–1468, 2007.

Part II
Filters On-Chip

Filtering has always been a very important topic in electronic design. The oldest forms of electronic filters were passive analog filters, using discrete resistors, capacitors and inductors and were unsuitable for on-chip integration. This changed with the first on-chip implementation of switched-capacitor filters in the early 1970s and the development of high performance, active on-chip filters from the second half of the decade on. Notwithstanding the inroads into this domain by submicron digital filters, analog on-chip filtering remains after more than 30 years of history a thriving domain of research and new architectures.

The first paper reviews the synthesis of low-sensitivity analog filters and proposes a design strategy, based on group-delay and pass-band ripple. Chebyshev II and Cauer approximations are found to be the best choices. A further sensitivity reduction is obtained by using a diminishing pass-band ripple approximation.

The following three papers discuss various aspects and architectures of continuous time filters. The first paper addresses the linearity limitations and the sensitivity to process variations of OTA-C filters. Linearization techniques and trade-offs are first discussed and a software-based calibration scheme is used to address the process sensitivity.

The second paper presents two different topologies of continuous-time filters, based on source follower circuits. They achieve a large linearity for a low overdrive, have low power consumption and require no common-mode feedback. Complex poles are synthesized using local positive feedback.

The third paper discusses in detail the architecture optimizations and circuit topologies of a wideband reconfigurable active-RC filter in order to achieve the high linearity and low noise requirements for home networking applications.

Instantaneously companding switched-capacitor filters are discussed in the next paper. They are advantageous over conventional automatic gain control techniques for signals with high peak-to-average-power ratio but are sensitive to DC offset in the OpAmp.

The last paper extends the notion of filters on-chip by co-integrating technologically compatible BAW resonators above the IC. Process dispersions and thermal drift of these resonators require tunable resonators and tuning circuitry.

Herman Casier

Chapter 6
Smart CMOS Current-Steering D/A-Converters for Embedded Applications

Martin Clara, Daniel Gruber, and Wolfgang Klatzer

Abstract The current-steering D/A-converter is the workhorse for the synthesis of high-resolution, wide-bandwidth analog signals, e.g. in the transmitter section of digital transceivers. Highly integrated systems require the implementation of such circuits in a nanometer CMOS technology together with analog and digital signal processing functions. Multi-mode operation additionally complicates the design task, since it is desired to minimize the circuit overhead in terms of silicon area and power consumption. "Smart" data converters make use of auxiliary analog and digital circuitry to enhance the linearity and to eventually tailor the converter architecture to the specific operating mode.

6.1 Introduction

Although in modern communication systems almost all signal processing is performed in the digital domain using powerful DSP's, the transmission medium, or channel, has an analog nature. Prior to transmission, the coded and modulated digital data stream must therefore be transformed into an analog signal using the on-chip D/A-converter, see Fig. 6.1. At the receiver-side, prior to digital demodulation, the received analog input signal is finally digitized by an embedded A/D-converter [1].

Demanding digital communication systems, for example DSL or BB-PLC, require a transmitter resolution of 12–14 b with analog bandwidths up to 50 MHz [2, 3]. For these applications the current-steering architecture is usually chosen, mainly because in advanced CMOS-technologies it can easily be integrated together with large portions of the digital signal processing chain, along with analog post-processing functions such as filtering and amplification of the synthesized analog output signal.

A special situation is the design of data converters for multiple operating modes with different bandwidth and linearity requirements. To achieve a minimum

M. Clara (✉), D. Gruber, and W. Klatzer
Infineon Technologies AG, Villach, Austria
e-mail: martin.clara@infineon.com

A.H.M. van Roermund et al. (eds.), *Analog Circuit Design: Smart Data Converters, Filters on Chip, Multimode Transmitters*, DOI 10.1007/978-90-481-3083-2_6,
© Springer Science+Business Media B.V. 2010

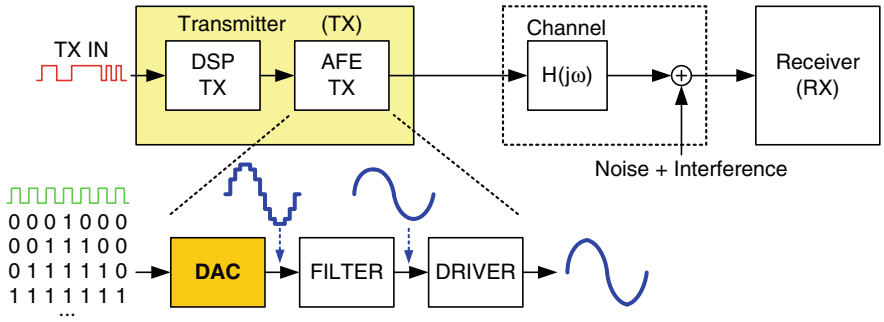


Fig. 6.1 Digital communication link

overhead in terms of power dissipation and silicon area, it may become necessary to change the clock rate, the digital signal format, the bias conditions, and sometimes even the converter architecture for different operating modes. Furthermore, in embedded applications the D/A-converter is just one building block among many others. In such a system, certain design constraints tend to receive increased attention, especially when multiple channels have to be integrated on the same silicon die:

1. Low power dissipation and small silicon area are of primary importance in SoC-implementations, since these parameters have the strongest impact on package cost. In multi-channel designs this problem is even more critical, because any increase in the power dissipation or silicon area of a single building block is immediately multiplied by the number of channels.
2. The number of power supplies that have to be made available in a system has a strong influence on overall cost and form factor. Each additional supply voltage requires a number of external components and a more complex PCB-routing. In the extreme case only a single voltage supply is available.
3. In SoC-design it is generally not possible to choose for every building block the optimum location on the silicon die such that the effect of thermal or stress gradients is minimized. Instead, the chip floorplan will try to optimize the silicon area, while maintaining reasonable signal flow and supply paths and separating sensitive blocks from potential disturbers. Thus, the performance of embedded converters must not depend on a special placement on the die.

Although the non-interruptive switching ("steering") of current sources is an open-loop operation and therefore intrinsically fast, the passively terminated current-steering DAC suffers from a limited voltage swing and does not interface naturally to typical on-chip analog building blocks, for example an op-amp based reconstruction filter. The use of an active output stage can solve this problem in low supply voltage designs, but this approach generally results in a power-bandwidth trade-off and must be carefully evaluated.

In the following sections three hardware examples of D/A-converters intended for flexible system integration are discussed. All modules have been implemented

in a 1P6M 130 nm CMOS technology, use only regular devices and operate from a single 1.5 V supply voltage. Two of the converters use an active output stage to maximize the on-chip voltage swing.

6.2 A Multimode ΣΔ-DAC

The block diagram of a ΣΔ-DAC for multiple bandwidth applications is shown in Fig. 6.2. The interpolation filter and the digital noiseshaper can be configured by register programming for different interpolation ratios and different noiseshaper structures. The clocks needed for the interpolation filter, the noiseshaper, the DEM-logic and the converter core are generated from an on-chip PLL-clock using a programmable clock divider. To reduce the static power dissipation in lower-bandwidth modes, the bias currents of the operational amplifier in the output stage can be programmed in a wide range and allow tailoring the amplifier performance to the requirements given by the converter and signal characteristics.

The structure of the reconfigurable digital noiseshaper is shown in Fig. 6.3. By making all coefficients a_i, b_i and g_1 programmable, it allows to implement a third-order NTF with in-band zero, or almost any other lower-order NTF. For example, by optimizing the coefficient set for a third-order noiseshaping function, a dynamic range of 81 dB in the digital domain for a low oversampling ratio of 6 can be achieved with a 6-b DAC-core. With a clock rate of 350 MHz the resulting analog bandwidth is 29.16 MHz. Conversely, by setting the coefficients a1, b1, b3, b4 and g1 to zero, a standard second-order noiseshaper can be implemented. In this case a digital dynamic range of over 90 dB for an oversampling ratio of 24 is easily possible. This corresponds, for example, to an analog bandwidth of 2.2 MHz using a clock rate of 106 MHz. Allowing for a comfortable analog design margin of about 1 b, the converter can thus be configured by software to a 12-b DAC having an analog bandwidth of almost 30 MHz, as well as a 14-b converter able to cover analog bandwidths in the lower single-digit MHz range.

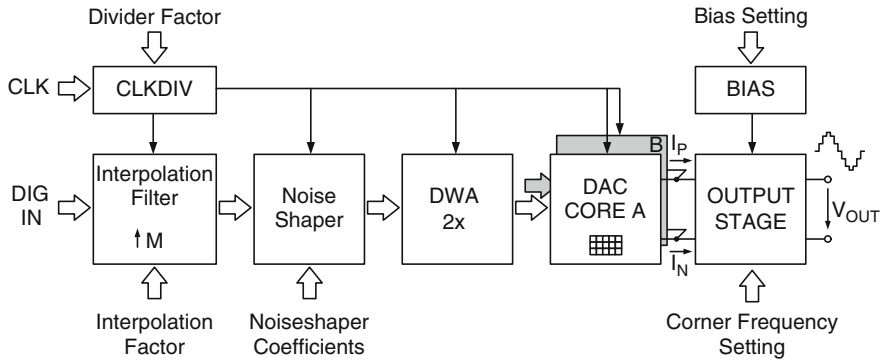


Fig. 6.2 ΣΔ-DAC for multiband operation

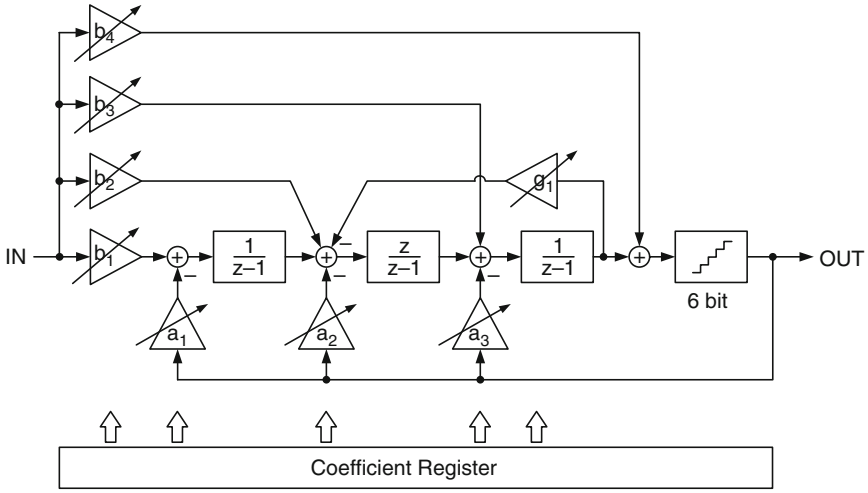


Fig. 6.3 Reconfigurable digital noiseshaper

Data Weighted Averaging (DWA) is a simple and thus power efficient DEM-algorithm with a first-order mismatch shaping property [4]. As a drawback, DWA-switching has a strong signal correlation, such that injected error charges associated with the switching of the current sources are very efficiently converted into harmonic distortion products [5]. A possible solution is the implementation of a RZ-architecture. The inherent 6 dB signal loss encountered in the classical half-clock reset-scheme, as well as the increased sensitivity toward clock jitter makes this approach however less desirable for low-noise systems. "Double Return-to-Zero", as described in [6], in principle solves the problem, but it leads to a relatively high switching activity resulting in a considerable power dissipation penalty at the target clock rate of 350 MHz. To reduce the amount of switching, an interleaved current cell architecture as shown in Fig. 6.4 is used [7]. Each DAC-element is represented by two current-cells (A + B). Each of the cells is alternately active for one clock cycle and reset during the following clock cycle. By summing the output current of the two cells, an NRZ output current is generated, while each individual current cell still experiences a full-clock RZ-behavior.

The analog section of the converter is shown in Fig. 6.5. It consists of two converter cores operated in interleaved mode providing a full-clock RZ for each current cell. The output currents are summed at the virtual ground nodes of the output stage and converted into a differential output voltage with the feedback impedance. Since in an interleaved current-source array the matching between the two sub-converters is very critical, the current cells of the two DAC-cores are arranged in a single matrix using interleaved placement for maximum proximity. Tight synchronization is achieved by symmetric clock trees [7].

Figure 6.6 shows a two-tone measurement result. A 25 MHz and 27 MHz tone, each at −7 dBFS, are sampled at 350 MHz using the third-order noiseshaper and

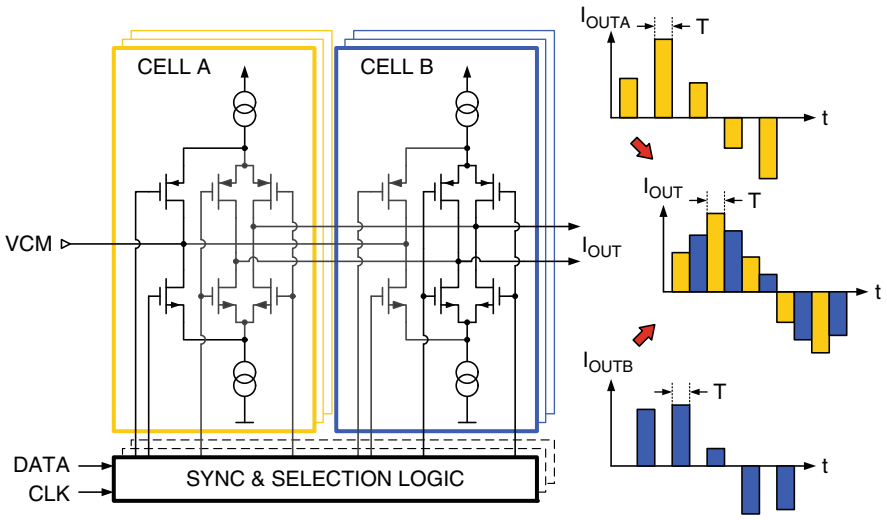


Fig. 6.4 Interleaved-RZ current-cell array

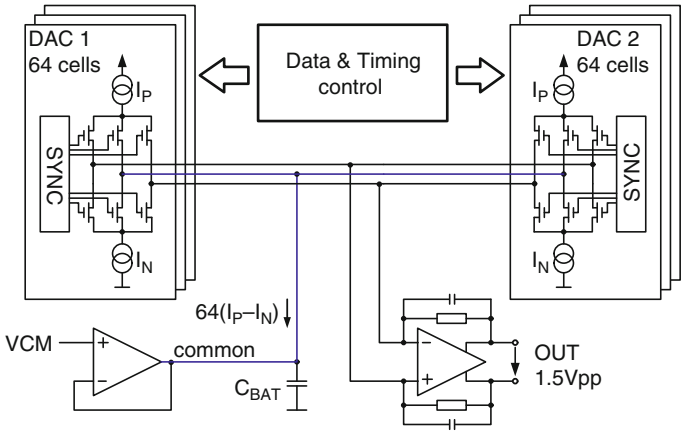


Fig. 6.5 Converter-core implementation

interleaved DWA. The IMD3 is at −76 dBc. With an OSR of 6 the converter achieves a DR of 73.4 dB for an analog bandwidth of 29.16 MHz. At the sampling rate of 350 MHz the complete converter module draws 62 mW from a single 1.5 V supply.

In Fig. 6.7 a 130 kHz single-tone signal sampled at 106 MHz using a second-order noiseshaper is shown. The linearity is limited by HD3 at −85 dBc. In this mode the converter achieves a DR of 86 dB for an analog bandwidth of 2.2 MHz (OSR = 24). Besides reconfiguring the digital part of the converter, also the bias currents in the output stage are reduced to 25% of their nominal value, resulting in a total power dissipation of only 19 mW from the 1.5 V supply.

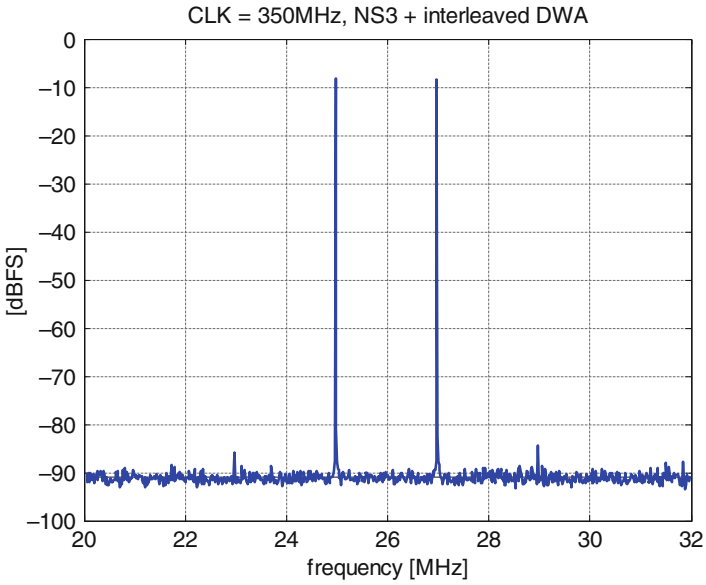


Fig. 6.6 Two-tone signal @ 350 MS/s

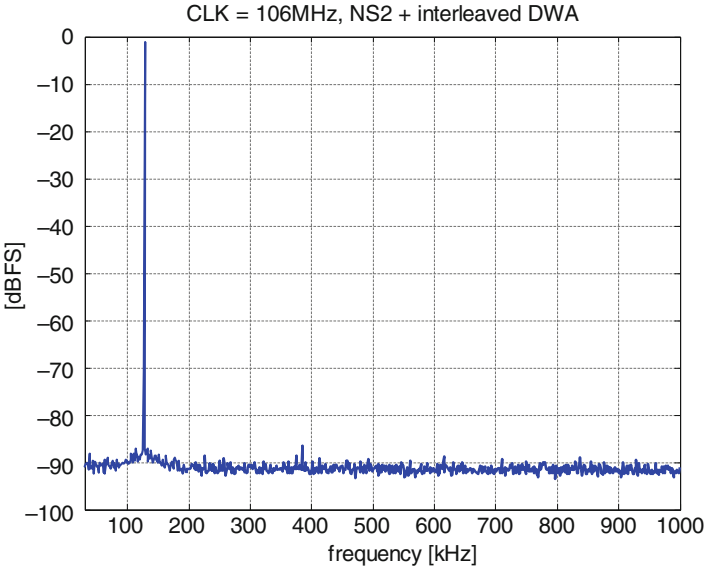


Fig. 6.7 Single-tone signal @ 106 MS/s

6.3 A 13-b 200 MS/s Background-Calibrated DAC

While DEM is an error averaging technique applied during data processing, calibration tries to achieve an accurate element value *before* the current cell is used for signal synthesis. Factory trimming is a one-time calibration method applied directly after the fabrication of the integrated device. For example, laser trimming of thin-film NiCr-resistors can achieve 16-b accuracy [8, 9]. However, factory calibration requires a special process and expensive trimming equipment. It is therefore only used for precision stand-alone devices.

Startup-calibration requires a dedicated calibration phase for the converter, e.g. at power-on. Either the single DAC-elements are compared one-by-one with a reference current and a small correction current is applied [10], or the complete converter characteristic is tracked using a dedicated measurement ADC and corrected with a global calibration DAC [11]. Since the converter cannot be used during the calibration phase, slow variations of temperature and bias conditions during normal operation typically cannot be corrected.

Dynamic current calibration in the background is based on dynamic current copying, originally introduced to achieve very accurate current mirror ratios [12]. The extension of this principle to complete current source arrays forming a current-steering D/A-converter was first described in [13]. By including adequate redundancy in the current-cell array, it allows uninterrupted converter operation, while its elements are periodically re-trimmed in the background.

The basic principle of the dynamic current calibration is shown in Fig. 6.8. The current-cell chosen for trimming in the background is replaced in the DAC-array by a redundant element and put into a feedback loop by closing the switches S_1 and S_2. The feedback forces the cell current to be equal to the reference current I_{REF}. When the network is settled, the correct gate-source voltage is stored on capacitor C_S, partly or totally represented by the gate-source capacitance of transistor M_V. Now switches S_1 and S_2 can be opened and the current-cell put back into normal

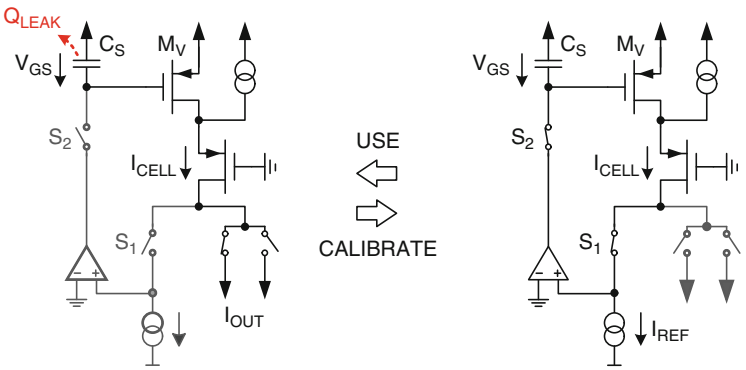


Fig. 6.8 Dynamic current calibration principle

operation. Because the charge stored on C_S will leak away over time, the calibration process has to be repeated periodically for each current-cell of the array in order to maintain the accuracy of the converter.

6.3.1 Converter Architecture

Figure 6.9 shows the architecture of the converter. The 13-b converter core is segmented into a 6-b unary MSB-array, a 2-b unary ULSB-array and a 5-b binary LLSB-array. Each segment displays adequate redundancy to allow uninterrupted data processing, while performing element-wise calibration in the background. Since the ULSB- and LLSB-arrays are calibrated as a whole, as described in Section 6.3.2, they are duplicated. Additionally, each segment contains an extra element. The converter uses on-chip load resistors to convert the signal current into a differential output voltage.

Typically, two additional problems must be solved in a segmented converter using dynamic background calibration. First, the lower segments have to be matched to the calibrated MSB-elements. Traditionally, this issue is solved by current division [10, 13] or with accurate current mirrors [14]. Another approach is described in the following section. The second problem is the generation of spurious tones in the output spectrum due to the periodicity of the calibration process. A possible solution is described in Section 6.3.3.

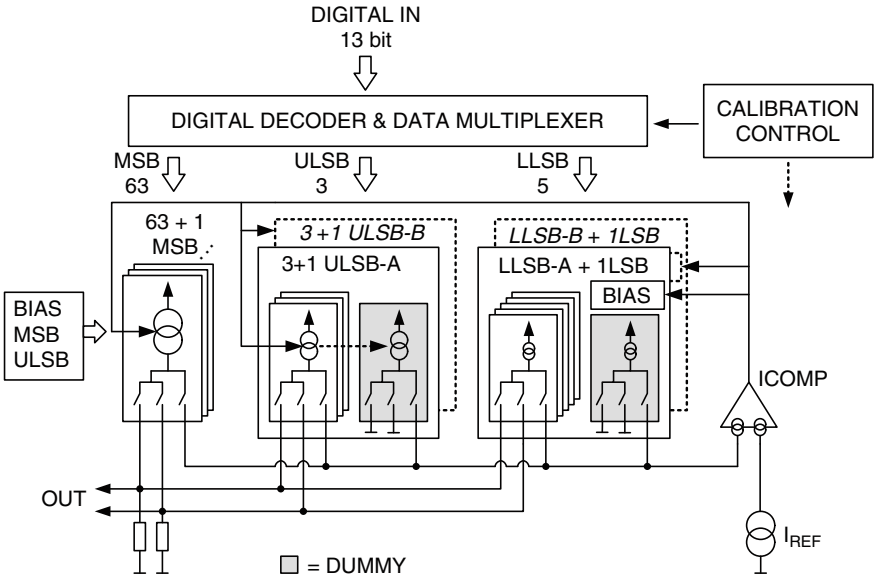


Fig. 6.9 Converter architecture

6.3.2 Segment Boundary Calibration

To solve the segmentation problem, we can also try to match the segment boundaries [15]. If the sum of all elements in one of the lower segments plus one unit current of that segment is equal to the unit current of the next higher segment, then the whole static characteristic of the converter is correct, provided that the matching within all lower segments is guaranteed by design.

Figure 6.10 shows the calibration loop for an MSB-cell. The subtraction of the cell current I_{MSB} from the reference current I_{REF} occurs at the low-impedance node A, i.e. at the source of cascode transistor N_0. At node B the current difference generates a voltage that is amplified and drives the gate of transistor M_V, such that $I_{MSB} - I_{REF} = 0$. Capacitor C_C stabilizes the feedback loop.

To match the ULSB-array to the MSB unit-cells, four identical ULSB-cells are switched in parallel for calibration, see Fig. 6.11. The feedback loop forces their

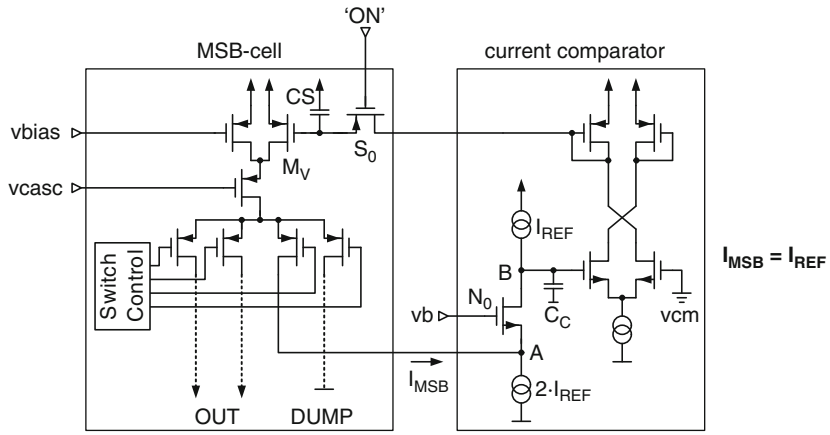


Fig. 6.10 MSB-cell calibration

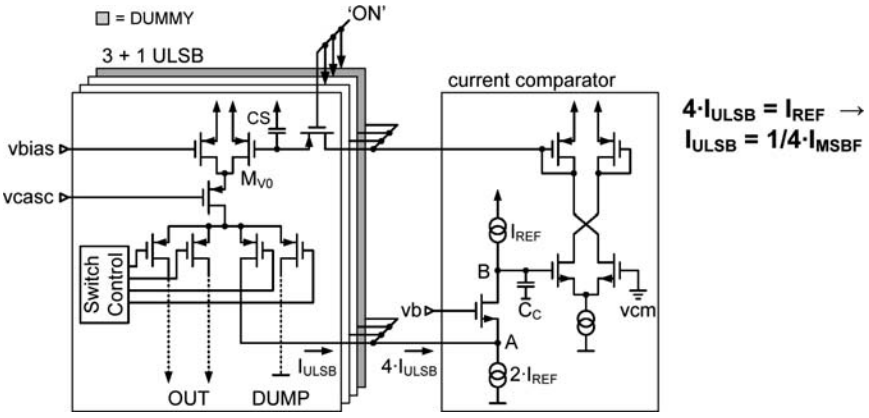


Fig. 6.11 ULSB-array calibration

sum to be identical to the reference current. This means that the unit ULSB-current is one fourth of the calibrated MSB-current, given sufficient intra-segment matching within the ULSB-array. The fourth ULSB-current is only used during calibration and dumped otherwise.

Next, the sum of the LLSB-array must be matched to the ULSB unit-current. The LLSB-array plus one extra LSB-current (only used during calibration), is summed to the current of three, previously trimmed, ULSB-currents. The LLSB-array is then adjusted with the bleeding transistor M_{VB} in the current mirror, such that the overall current sum is equal to the reference current (see Fig. 6.12). This is equivalent to making the sum of the LLSB-array equal to the ULSB unit-current minus a single LSB-current.

The effect of the different calibration steps can be observed in the THD-plot of Fig. 6.13. The THD of a synthesized full-scale 1 MHz sine wave is monitored as a

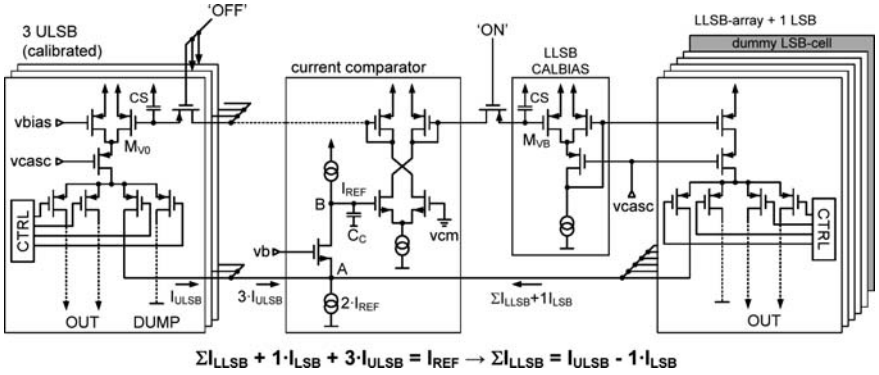


Fig. 6.12 LLSB-array calibration

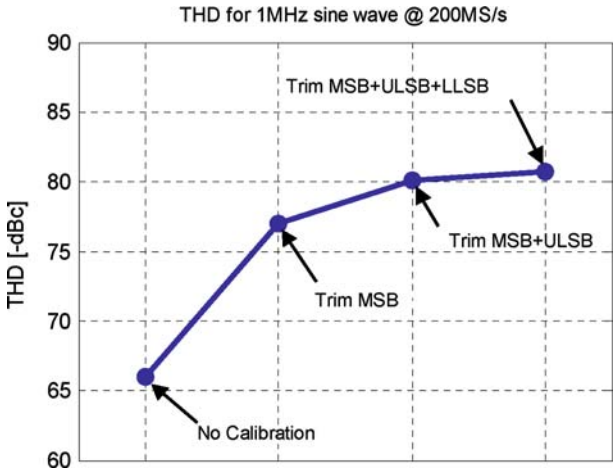


Fig. 6.13 THD as function of the calibration depth

function of the segmented calibration depth. Without calibration a typical THD of 66 dB, limited by the mismatch of the current cells, is measured. By calibrating the MSB-cells only, the THD rises to 76 dB. Including the ULSB-array in the calibration loop improves the THD to 80 dB, whilst the full calibration loop, including also the LLSB-array, pushes the THD to 81 dB.

6.3.3 Randomization of the Calibration Period

The periodicity of the calibration process generates spurious tones in the output spectrum at multiples of the fundamental refresh frequency. The main reason for this behavior is dynamic mismatch related to the switching of the current cells into calibration mode and back. A current-cell architecture with "backside" access for measurement (also called "floating" current source [16]) allows trimming during data processing and doesn't require redundant current cells. Because the output-side switching into calibration mode and back is eliminated in this scheme, it results in a considerable attenuation of the calibration spurs [14, 16]. However, the required current folding at the output increases the static power dissipation of the converter.

Another possibility is to eliminate the periodicity of the calibration process [15]. The calibration slot length, i.e. the time assigned to the calibration of a single DAC-element, can be varied between a minimum and maximum duration without compromising the converter accuracy. The minimum calibration slot length is given by the settling time required by the feedback loop, while the maximum length is limited by the total leakage current that detunes the voltage on the storage capacitor. Shown in Fig. 6.14 is the generation of a random calibration slot length between these two extreme values. A random number generator (LFSR) delivers a random number N_{VAR}, which is added to a fixed number N_{FIX}, the latter indicating the minimum calibration time. The sum of both, N_{CAL}, is loaded into a down-counter. When the counter expires, the current calibration slot is terminated and a new biased random number is generated for the next current cell.

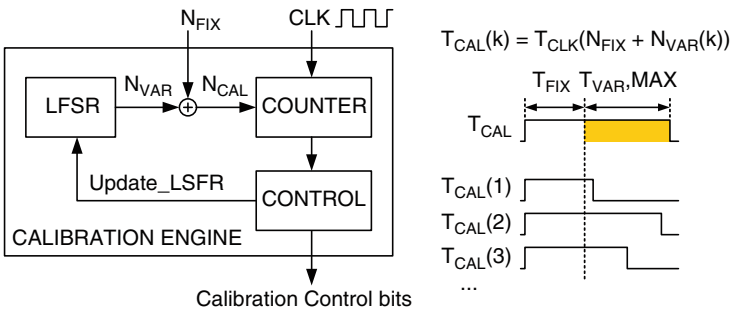


Fig. 6.14 Generation of random calibration slot length

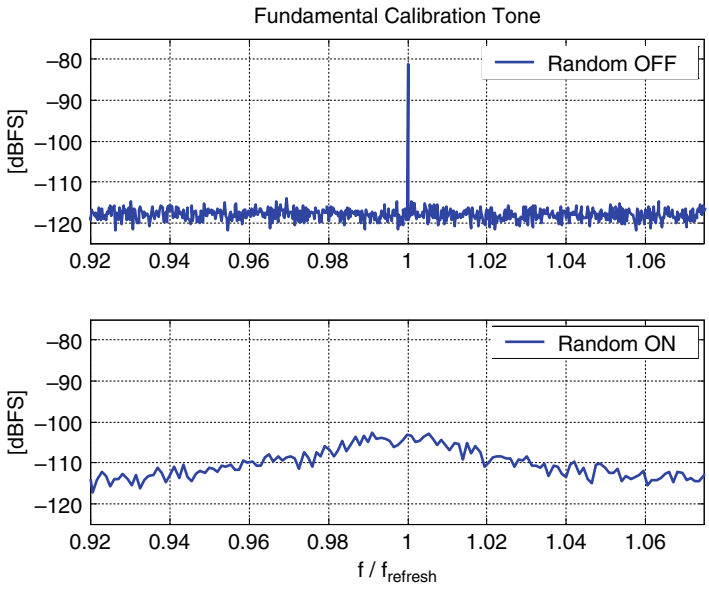


Fig. 6.15 Effect of calibration timing randomization

Figure 6.15 shows the effect of the randomization of the calibration slot length. With a fixed calibration period, the spurious tone at the fundamental refresh rate is at around −80 dBFS. When the randomization is activated, the energy residing in the calibration spur is spread over a wider bandwidth, eventually merging with the system noise floor. In this example, a reduction of >20 dB of the maximum calibration tone magnitude is possible.

The output spectrum of a 10 MHz sine-wave sampled at 200 MS/s is shown in Fig. 6.16. The nonlinearity of the converter is dominated by the second harmonic distortion at −73.6 dBc, while the third harmonic distortion is at −78 dBc.

6.3.4 Low-Bandwidth High-Resolution Mode

A unary array has inherently more static and dynamic homogeneity compared to a segmented converter. This property usually results in a better linearity and resolution, especially when processing multi-tone signals. In order to exploit this advantage in lower-bandwidth modes, the MSB-segment of the DAC can be operated as an oversampled 6-b unary converter together with a compatible digital noiseshaper, while the lower segments are not used [17].

Shown in Fig. 6.17 is the structure of such a dual-mode converter. Instead of the full segmented converter with 13-b input, the 6-b MSB-array with a second-order digital noiseshaper is used. To avoid DC-offset at the DAC-output in this operating

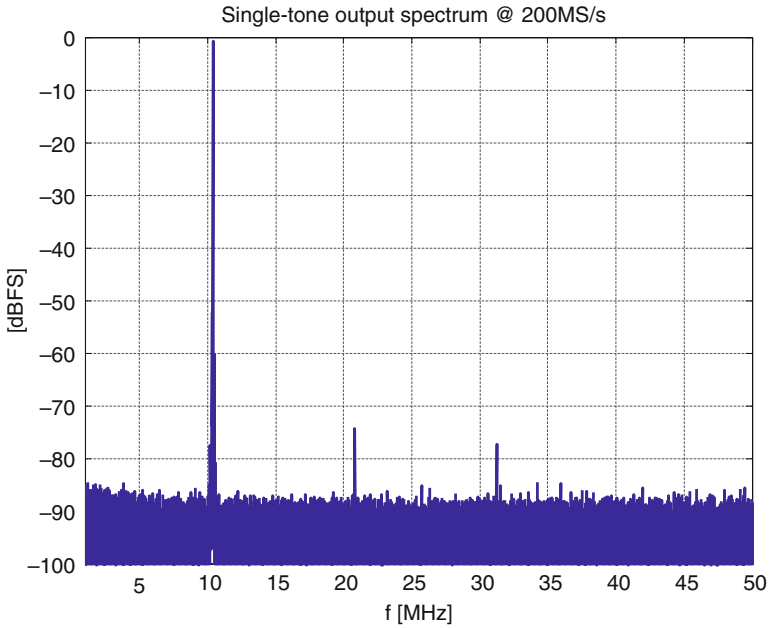


Fig. 6.16 10 MHz single-tone signal sampled at 200 MS/s

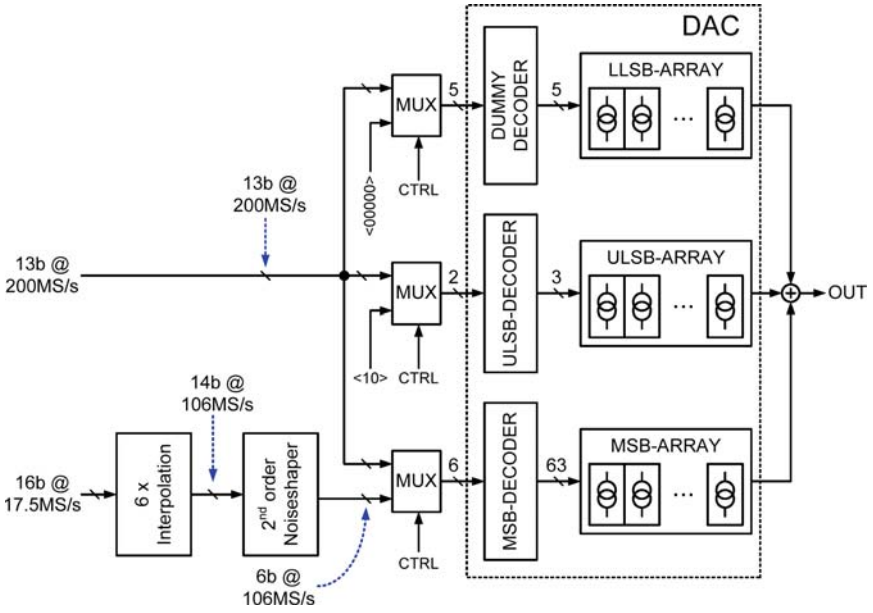


Fig. 6.17 Multimode converter

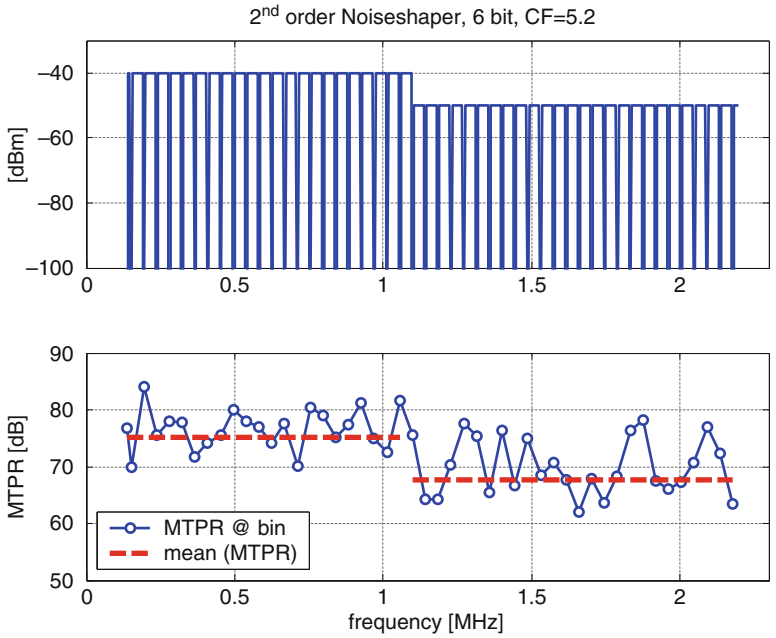


Fig. 6.18 Noiseshaped 6-b DMT-signal @ 106 MS/s

mode, the ULSB and LLSB-array are brought into a balanced position. Figure 6.18 shows the Missing-Tone-Power-Ratio (MTPR) of an ADSL2 + downstream signal with Peak-to-Average-Ratio (PAR) of 5.2 synthesized at 106 MS/s. In this design the achievable MTPR of such low-bandwidth signals using the oversampled and noiseshaped MSB-segment is on average 2–3 dB better than with the full segmented converter.

6.4 A 13-b 50 MHz Bandwidth DAC with Active Output Stage

In low supply voltage designs, the passive termination of the current-cell array using a pair of resistors connected to the ground rail limits the achievable output swing of the DAC and is not directly compatible with typical on-chip building blocks connected to the converter, e.g. an op-amp based filter. With an active transimpedance stage the output voltage swing of the converter can be maximized, albeit at the cost of increased power dissipation. By optimizing the feedback impedance, the required slew-rate and thus the power dissipated in the amplifier can be traded off, to first order, with the achievable attenuation at the maximum signal frequency of interest [18]. In some systems the resulting LHP-pole in the transfer function of the output stage can be included in the overall reconstruction filter characteristic. Another possibility is the use of a digital preemphasis filter.

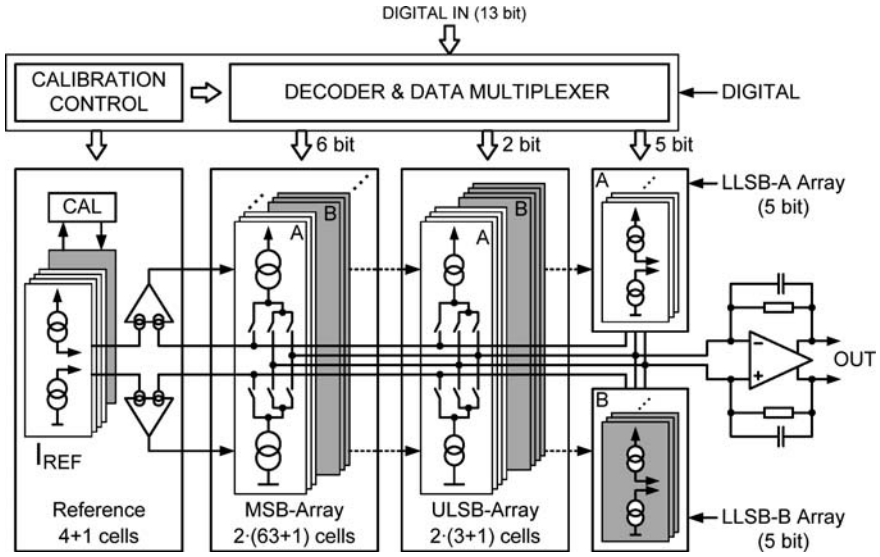


Fig. 6.19 Converter architecture

6.4.1 Converter Architecture

The converter architecture is shown in Fig. 6.19 [19]. The DAC-core is a 13-b current-steering array consisting of complementary current cells, segmented into a 6-b unary MSB-array, a 2-b unary ULSB-array and a 5-b binary LLSB-array. In the MSB- and ULSB-array interleaved current cells (A + B) are implemented. The output stage can drive a 1 kΩ differential on-chip load with a differential voltage swing of 1.5 Vpp. It is optimized for a 50 MHz analog bandwidth with a maximum excess attenuation of 1 dB.

6.4.2 Direct Segment Calibration

Direct calibration of current cells in differently weighted segments requires the generation of different reference currents with an accurate ratio. This can be achieved with the multi-level reference cell shown in Fig. 6.20. Five unit reference currents are trimmed in a separate background calibration loop using a "root" reference current of ULSB-size. Four reference cells are always available for calibration of the main DAC-elements: the MSB-cells are trimmed using all four reference currents, while for the ULSB-cells a single reference current is sufficient. For the LLSB-array a boundary trimming is performed, also using a single reference current. The status bit MSB_CAL indicates to the control logic the value of the reference current that must be provided for the main-DAC calibration during the actual calibration slot.

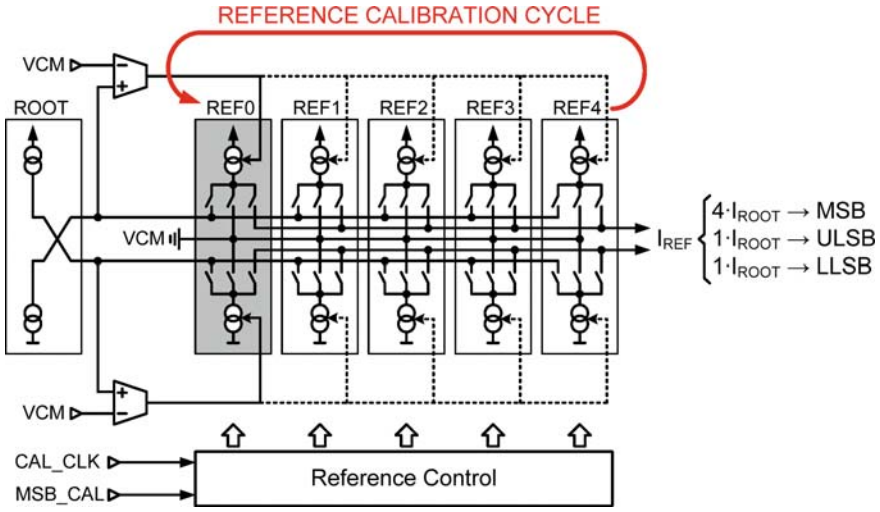


Fig. 6.20 Multi-level reference cell

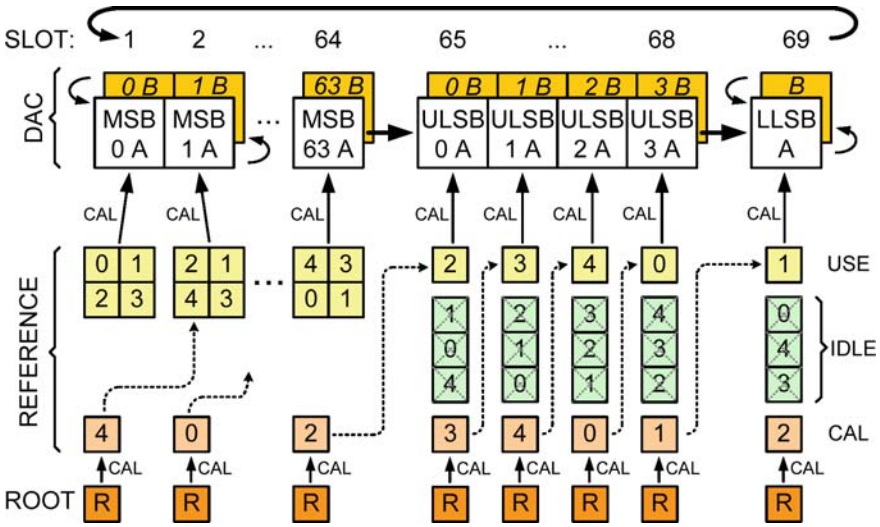


Fig. 6.21 Calibration cycle

The complete calibration cycle is shown in Fig. 6.21. In the first 64 slots the MSB-cells are trimmed. The reference current comes from four changing reference cells with the fifth cell always in calibration mode. Since the ULSB-cells and the LLSB-array require only one reference current, three reference cells remain idle during ULSB- and LLSB-calibration, while again always one reference cell is calibrated. The reference calibration cycle runs in parallel to the main-DAC calibration, reusing the available random timing generation.

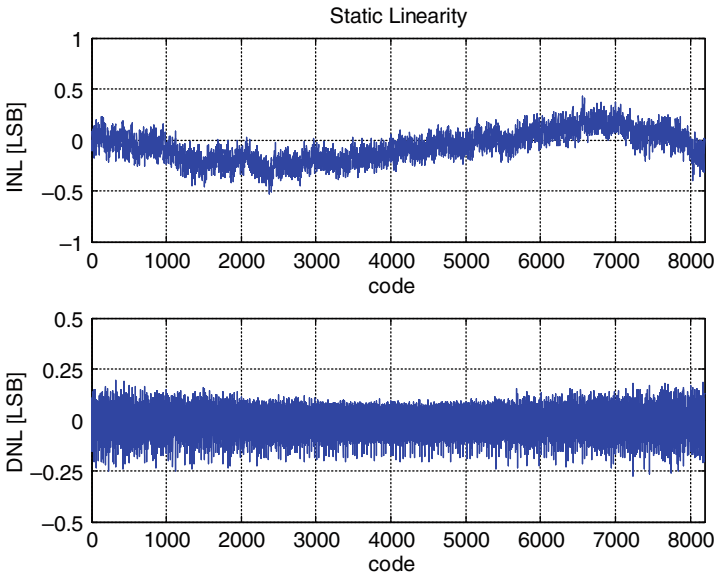


Fig. 6.22 Static linearity

The measured static linearity of the calibrated converter is shown in Fig. 6.22. DNL and INL are $-0.27/+0.19$ LSB and $-0.53/+0.42$ LSB respectively.

The converter can be operated in Nyquist-mode with a clock rate up to 130 MHz. Allowing 1 dB excess attenuation the achievable analog bandwidth is 50 MHz for a total power dissipation of 53 mW. However, the order of the reconstruction filter can be considerably relaxed by operating the converter with a sampling rate of 300 MHz. In this case a 3× digital interpolation filter is used, again resulting in an analog bandwidth of 50 MHz. Although the power dissipation of the converter at 300 MS/s rises to 73 mW, the reconstruction filter will be much simpler and way less power hungry. For both sampling rates the active transimpedance output stage draws 30 mW from the 1.5 V supply.

The SFDR for a single-tone signal and the IMD3 for a two-tone frequency sweep are shown in Fig. 6.23. At 300 MS/s the SFDR for signal frequencies larger than 10 MHz is limited by crosstalk coming from the 3× digital interpolation filter. The dotted line indicates the achievable full-Nyquist SFDR after (theoretical) elimination of the spurious tones generated by digital injection. With this improvement the SFDR would always remain above 70 dB over the whole 50 MHz signal bandwidth.

A four-band multitone signal with a bandwidth of 43 MHz synthesized at 300 MS/s is shown in Fig. 6.24. The Peak-to-Average-Ratio (PAR) of the signal is 5.4 with an in-band carrier spacing of 50 kHz. The worst-case MTPR is better than 62 dB and the MBPR better than 65 dB.

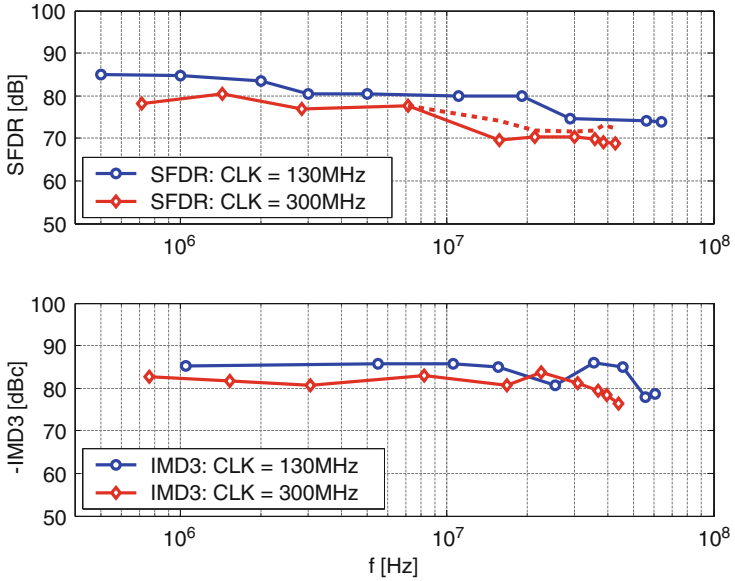


Fig. 6.23 SFDR and IMD3 for 130 MS/s and 300 MS/s

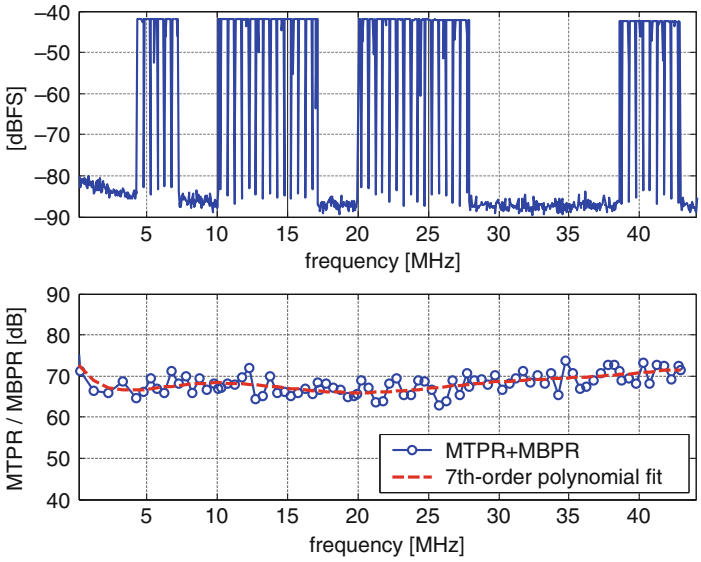


Fig. 6.24 43 MHz multiband-DMT signal with CF = 5.4 @ 300 MS/s

6.5 Conclusions

D/A-converters embedded into digital transceivers make extensive use of analog and digital techniques to enhance the linearity and increase the flexibility toward multi-mode operation. The continuous reduction of the output voltage range as a result of the supply voltage scaling in newer technologies can be circumvented by using an active output stage, although resulting in a trade-off between power dissipation and achievable analog bandwidth. Since digital circuits are becoming "cheaper" with the migration to finer CMOS-technologies, it is expected that the amount of "digital assistance" will be steadily increasing in future designs.

References

1. R. Gaggl, "Design of Embedded CMOS A/D-Converters for Communication Systems", PhD-Dissertation, Graz University of Technology, 2009.
2. P. Golden, H. Dedieu, K. Jacobsen (editors), "Implementation and Applications of DSL Technology", ISBN 0-8493-3423-3, CRC-Press, Taylor & Francis Group, 2008.
3. M. Delgado-Restituto, J. Ruiz-Amaya, J. M. de la Rosa, J. F. Fernández-Bootello, L. Díez, R. del Río, Á. Rodríguez-Vázquez, "An Embedded 12-bit 80 MS/s A/D/A Interface for Power-Line Communications in 0.13 μm Pure Digital CMOS Technology", Proceedings of the 2005 International Symposium on Circuits and Systems, Vol. 5, pp. 4626–4629, May 2005.
4. R. Baird and T. Fiez, "Improved ΣΔ DAC Linearity Using Data Weighted Averaging", Proceedings of the 1995 International Symposium on Circuits and Systems, Vol. 1, pp. I-13–I-16, May 1995.
5. M. Clara, A. Wiesbauer, W. Klatzer, "Nonlinear Distortion in Current-Steering D/A-Converters due to Asymmetrical Switching Errors", Proceedings of the 2004 International Symposium on Circuits and Systems, Vol. 1, pp. I-285–I-288, May 2004.
6. R. Adams, K. Q. Nguyen, K. Sweetland, "A 113-dB SNR Oversampling DAC with Segmented Noise-shaped Scrambling", IEEE JSSC, Vol. 33, No. 12, pp. 1871–1878, 1998.
7. M. Clara, W. Klatzer, A. Wiesbauer, D. Straeussnigg, "A 350 MHz low-OSR ΣΔ Current-Steering DAC with Active Termination in 0.13 μm CMOS", IEEE International Solid-State Circuits Conference 2005, Digest of Technical Papers, Vol. 48, pp. 118–119, 2005.
8. T. Guy, L. Trythall, A. Brodersen; "A Sixteen-bit Monolithic Bipolar DAC", IEEE JSSC, Vol. 17, No. 6, pp. 1127–1132, 1982.
9. J.R. Naylor, "A Complete High-speed Voltage Output 16-bit Monolithic DAC", IEEE JSSC, Vol. 18, No. 6, pp. 729–735, 1982.
10. W. Schofield, D. Mercer, L. St. Onge, "A 16 b 400 MS/s DAC with < -80 dBc IMD to 300 MHz and < -160 dBm/Hz Noise Power Spectral Density", IEEE International Solid-State Circuits Conference 2003, Digest of Technical Papers; pp. 126–129, 2003.
11. Y. Cong, R. L. Geiger, "A 1.5-V 14-bit 100-MS/s self-calibrated DAC", IEEE JSSC, Vol. 38, No. 12, pp. 2051–2060, 2003.
12. G. Wegmann, E. Vittoz, "Very Accurate Dynamic Current Mirrors", Electronics Letters, Vol. 25, pp. 644–646, 1989.
13. W. Groeneveld, H. Schouwenaars, H. Termeer; "A Self Calibration Technique for Monolithic High-Resolution D/A Converters", IEEE Journal of Solid-State Circuits, Vol. 24, No. 6, pp. 1517–1522, 1989.
14. Q. Huang, P.A. Francese, C. Martelli, J. Nielsen; "A 200 MS/s 14b 97 mW DAC in 0.18 μm CMOS", IEEE International Solid-State Circuits Conference 2004, Digest of Technical Papers, pp. 364–532, 2004.

15. M. Clara, W. Klatzer, B. Seger, A. Di Giandomenico, L. Gori, “A 1.5 V 200 MS/s 13b 25 mW DAC with Randomized Nested Background Calibration in 0.13 μm CMOS”, IEEE International Solid-State Circuits Conference 2007, Digest of Technical Papers, pp. 250–251, 2007.
16. A. Bugeja, B.-S. Song, “A Self-Trimming 14-b 100-MS/s CMOS DAC”, IEEE Journal of Solid-State Circuits, Vol. 34, No. 12, pp. 1841–1852, 2000.
17. M. Clara, J. Hauptmann, “Digital-to-Analog Converter Arrangement with an Array of Unary Digital-to-Analog Converting Elements Useable for Different Signal Types”, US Patent No. 6,831,581, issued Dec. 14, 2004.
18. M. Clara, “High-Performance CMOS Current-Steering D/A-Converters for Digital Transceivers”, PhD-Dissertation, Graz University of Technology, 2009.
19. M. Clara, W. Klatzer, D. Gruber, A. Marak, B. Seger, W. Pribyl, “A 1.5 V 13 bit 130–300 MS/s Self-calibrated DAC with Active Output Stage and 50 MHz Signal Bandwidth in 0.13 μm CMOS”, Proceedings of the 34th European Solid-State Circuits Conference, pp. 262–265, 2008.

Chapter 7
Synthesis of Low-Sensitivity Analog Filters

Lars Wanhammar

Abstract Doubly resistively terminated *LC* filters are optimal from an element sensitivity point of view and are therefore used as reference filter for high-performance active filters. The later inherits the sensitivity properties of the *LC* filter. Hence it is important to design the reference filter to have minimal element sensitivity. In this paper, we first review the mechanism for the low sensitivity and give an upper bound on the deviation in the passband attenuation. Next we compare classical lowpass approximations with respect to their influence on the sensitivity and propose the use of diminishing ripple in the passband to further reduce the sensitivity. Finally, we propose a design strategy for doubly resistively terminated *LC* filters with low sensitivity.

7.1 Introduction

Passive *LC* filters belongs to the oldest implementation technologies, but still play an important role since they are being used in large volumes and are used as prototypes for the design of advanced frequency selective filters. A drawback with *LC* filters is that it is difficult to integrate resistors and coils with sufficiently high quality in integrated circuit technology. *LC* filters are for this reason not well suited for systems that are implemented in an integrated circuit.

A more important aspect is that *LC* filters are used as basis for realizing high-performance frequency selective filters. This is the case for mechanical, active, discrete-time, and *SC* filters as well as for digital filters. Examples of such methods are: immitance simulation using generalized immitance converters and inverters [8], Gorski-Popiel's [3, 8], Bruton's [1, 8], wave active filters [8], and topological simulations methods like leapfrog structures [2,8]. The main reason is that the magnitude function for a well-designed *LC* filter has low sensitivity in the passband for variations in the element values.

L. Wanhammar (✉)
Department of Electrical Engineering, Linköping University, Sweden
e-mail: larsw@isy.liu.se

A.H.M. van Roermund et al. (eds.), *Analog Circuit Design: Smart Data Converters, Filters on Chip, Multimode Transmitters*, DOI 10.1007/978-90-481-3083-2_7,
© Springer Science+Business Media B.V. 2010

It is important that the reference filter is designed for low sensitivity, since the active filter inherits its sensitivity properties. In fact, the sensitivity properties of the reference filter become a "lower" bound for the active filter.

7.2 Passive Filters

To explain the good sensitivity properties of correctly designed *LC* filters we first consider the power transferred from the source to the load in the circuit shown in Fig. 7.1. We assume that the input signal is, where V_1 is the effective value. The dissipated power in the load is $P = \mathrm{Re}\{V_2 I_2^*\} = \mathrm{Re}\{|V_2|^2/Z_2^*\}$, where V_2 is the r.m.s. value. The maximum power that can be transferred to the load impedance Z_2, in the circuit shown in Fig. 7.1, is

$$P_{2\,\max} = \frac{|V_1\ |^2}{4R_1} \tag{7.1}$$

and it is attained when $Z_2 = Z_1{}^*$ where $R_1 = \mathrm{Re}\{Z_1\}$ and $R_2 = \mathrm{Re}\{Z_2\}$. In the case of $\mathrm{Z}_1 = R_1 \neq Z_2 = R_2$, a transformer can be used to match the load to the source. Maximum power is transferred to the load if a transformer with the turns ratio n:1 is placed between the source and load resistors.

The input impedance to the primary side of the transformer is $n^2 R_2$ and the maximum power transfer occur if we select $R_1 = n^2 R_2$. In this case is equal amount of power dissipated in the resistor R_1 and R_2 and the ratio of the output and input voltages is 0.5.

7.2.1 Doubly Resistively Terminated Lossless Networks

Consider the doubly resistively terminated network in Fig. 7.2, which is lossless, i.e., it dissipates no power. A lossless reciprocal network can be realized by using only lossless circuit elements, e.g., inductors, capacitors, transformers, gyrators, and lossless transmission lines. Although these filters are often referred to as *LC* filters. The ratio of the output power and the maximal output power is

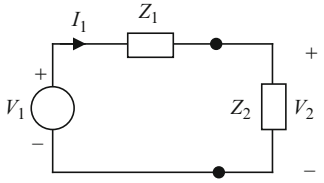


Fig. 7.1 Maximum power transfer if $\mathrm{Z}_2 = \mathrm{Z}_1{}^*$

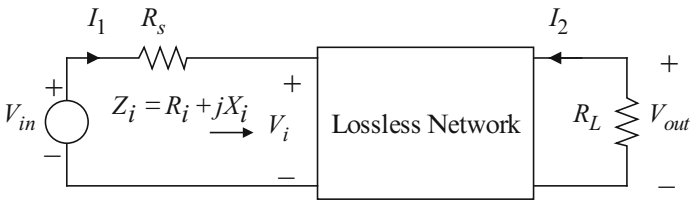


Fig. 7.2 Doubly resistively terminated reactance network

$$\frac{P_{out}}{P_{out\,\max}} = \frac{4R_s}{R_L}\frac{|V_{out}(j\omega)|^2}{|V_{in}(j\omega)|^2} \leq 1 \tag{7.2}$$

where $V_{\text{in}}(j\omega)$ is the r.m.s. value of the sinusoidal input signal. An important observation is that the power that the signal source can deliver to the load is limited. The upper bound for the maximal power transfer is the base for the design of filter structures with low element sensitivity.

We define the frequency response as the ratio between input and output voltages, i.e., the relation between signal quantities and corresponding physical signal carrier, according to

$$H(j\omega) = \sqrt{\frac{4R_s}{R_L}}\left|\frac{V_{out}(j\omega)}{V_{in}(j\omega)}\right| \tag{7.3}$$

Hence, the magnitude response is bounded from above according to

$$|H(j\omega)|^2 \leq 1 \tag{7.4}$$

7.2.2 Reflection Function

Note that the signal source shown in Fig. 7.2 does not deliver maximum power for all frequencies since the input impedance to the reactance network is frequency dependent. This can be interpreted as a part of the maximum available power is reflected back to the source. The relationship between the power that is absorbed in R_L and the power, which is reflected back to the source, can be derived in the following way.

The input impedance to the *LC* network is $Z_i(j\omega) = R_i(j\omega) + jX_i(j\omega)$. Since the reactance network is lossless, the power into the network will be absorbed in R_L, i.e.,

$$|I_1(j\omega)|^2 R_i(\omega) = \frac{|V_{out}(j\omega)|^2}{R_L} \tag{7.5}$$

Furthermore, we have

$$I_1(j\omega) = \frac{V_{in}(j\omega)}{R_s + Z_i(j\omega)} \tag{7.6}$$

After some simplifications we obtain

$$\frac{4R_s}{R_L}\,|H(j\omega)\,|^2 = 1 - \left|\frac{Z_i - R_s}{Z_i + R_s}\right|^2 \tag{7.7}$$

and Feldtkeller's Equation

$$\frac{4R_s}{R_L}\,|H(j\omega)\,|^2 = 1 - |\rho(j\omega)|^2 \tag{7.8}$$

where $\rho(j\omega)$ is the *reflection function* for port 1. The reflection function for port 2 is defined analogously.

7.2.3 Sensitivity

A measure of sensitivity is relative sensitivity of the magnitude function

$$S_x^{|H(j\omega)|} = \frac{x}{|H(j\omega|}\frac{\partial\,|H(j\omega)|}{\partial x} \tag{7.9}$$

It is difficult to find a simple and good measure of how the attenuation changes, when several circuit element vary at the same time. The reason for this is that the influence of errors in different element values interacts. In fact, for a doubly resistively terminated reactance network we will demonstrate that they tend to cancel. We shall therefore use and interpret sensitivity measures according to Eq. 7.9 with care. It is very difficult to compare different filter structures in a fair way.

7.2.4 Passband Sensitivity

The sensitivity of for example the magnitude function with respect to a circuit element, x, is a function of the angular frequency. The sensitivity in the passband can be determined from the derivative of the Feldtkeller's equation with respect to an arbitrary circuit element, x. We get

$$S_x^{|H(j\omega)|} = -\frac{R_L}{4R_s}\left|\frac{\rho(j\omega)}{H(j\omega}\right| S_x^{|\rho(j\omega)|} \tag{7.10}$$

For a doubly resistively terminated *LC* filter we have

$$|\rho_1(j\omega)| = |\rho_2(j\omega)| = \sqrt{10^{0.1A_{\max}} - 1} \tag{7.11}$$

where A_{max} is the acceptable ripple in the passband.

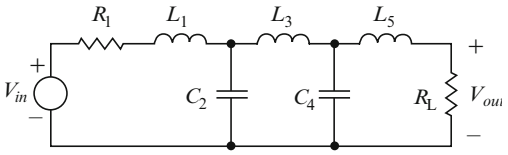


Fig. 7.3 Doubly resistively terminated LC ladder

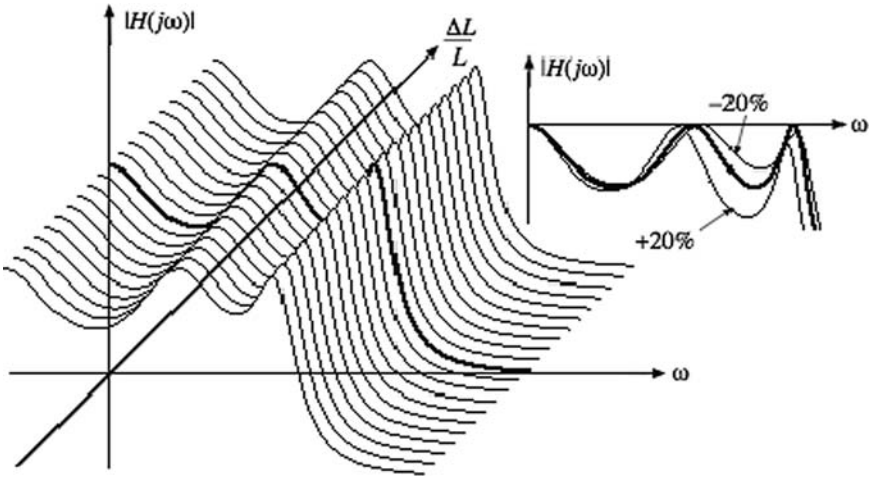


Fig. 7.4 Variation in the magnitude function due to ±20% variation in L_1 or L_5

Fettweis showed (1960) that the sensitivity becomes minimal if the filter is designed for maximum power transfer at a number of angular frequencies in the passband. At these angular frequencies, the reflection function $\rho(j\omega)$ is zero, since $Z_{in} = R_s$. The sensitivity at these frequencies is therefore, according to Eq. 7.10, zero. If A_{max} is small, both the reflection coefficient, according to Eq. 7.11, and the magnitude of the reflection function $|\rho(j\omega)|$, according to Eq. 7.8, will be small. Hence, the sensitivity will be small. If the ripple is increased in the passband, the sensitivity is also increased.

That a doubly resistively terminated *LC* filter has low element sensitivity can also be realized through the following reasoning. Irrespective of if the element value is increased or decreased from its nominal value, P_{out} will decrease, since $P_{out} = P_{out\ max}$ for the nominal element value. Since the derivative is zero where the function has a maximum, i.e., $\frac{\partial P_{out}}{\partial x} = 0$ for $\omega = \omega_k$ with nominal element values. If there are many angular frequencies, ω_k, with maximal power transfer in the passband, the sensitivity will be low throughout the passband. This line of reasoning is referred to as Fettweis-Orchard's argument [5, 8].

Example 7.1. Figures 7.4 through 7.8 show the sensitivity of the magnitude response for errors in the element values in the doubly resistively terminated ladder network shown in Fig. 7.3. The filter is a Chebyshev I filter with $A_{max} = 3\,\text{dB}$. The filter has been chosen with unusually large ripple in the passband to clearly demonstrate the sensitivity for errors in the element values.

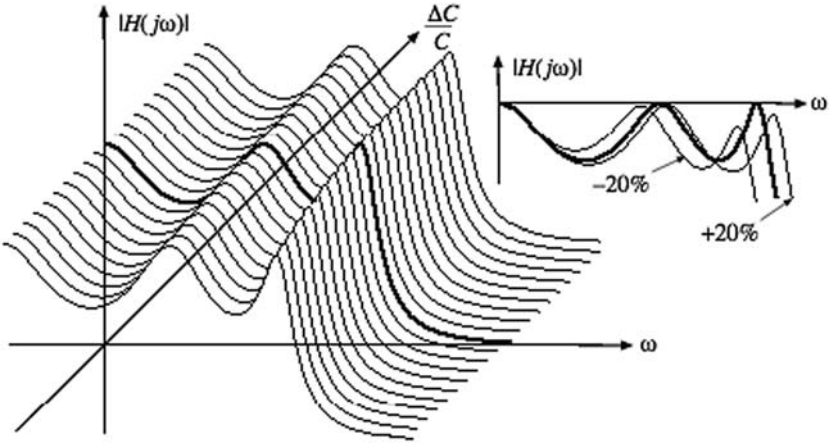


Fig. 7.5 Variation of the magnitude function due to ±20% variation in C_2 or C_4

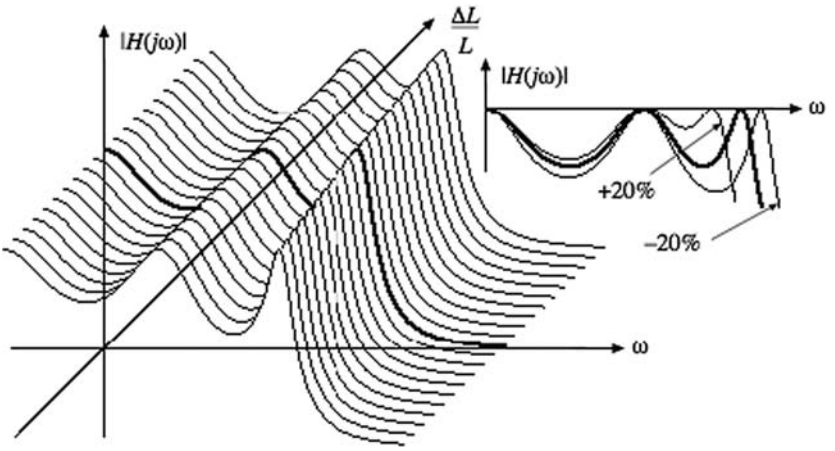


Fig. 7.6 Variation of the magnitude function due to ±20% variation in L_3

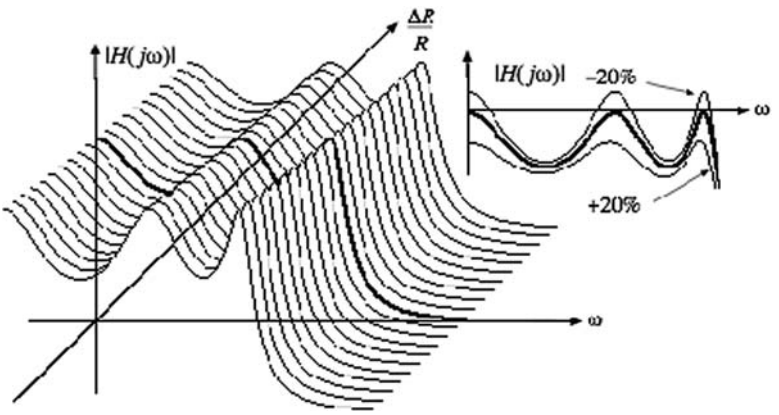


Fig. 7.7 Variation of the magnitude function due to ±20% variation in R_s

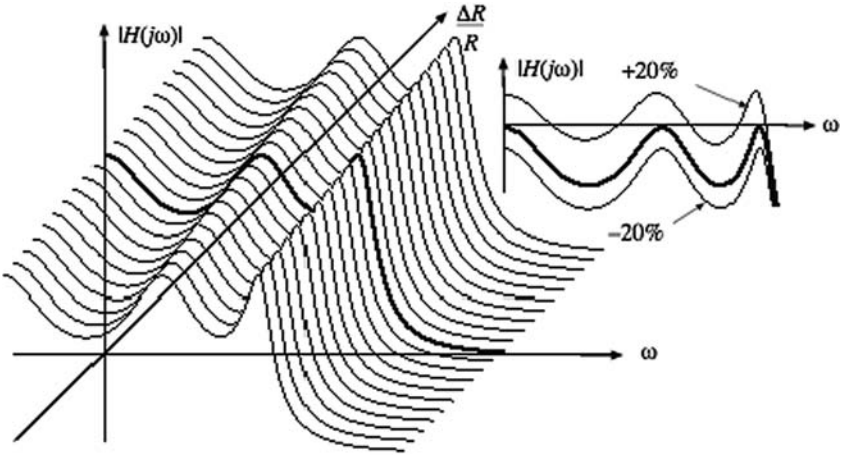
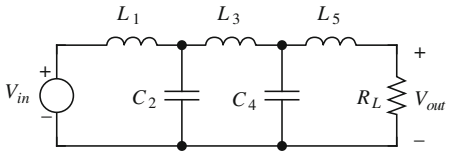


Fig. 7.8 Variation of the magnitude function due to ±20% variation in R_L

Fig. 7.9 Singly resistively terminated LC ladder



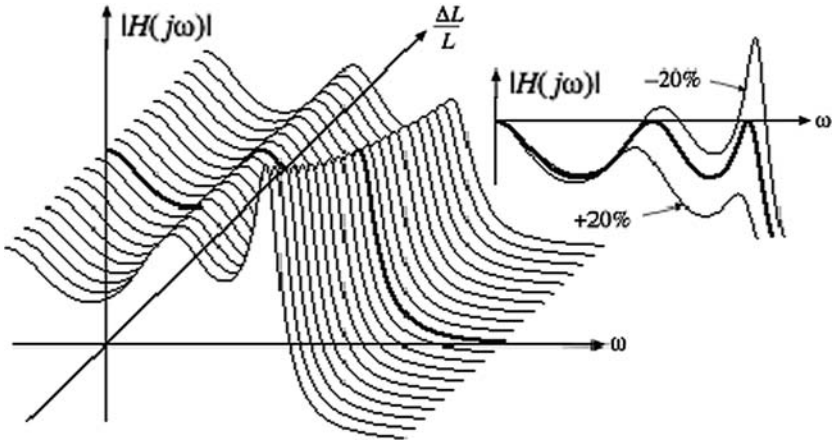


Fig. 7.10 Variation of the magnitude function due to ±20% variation in L_1

The element values are varied ±20% around the nominal values. Note that this variation is extremely large and is used only to clearly demonstrate the difference in element sensitivity.

Example 7.2. Figures 7.10 through 7.15 compare the sensitivity for errors in the element values in the singly resistively terminated ladder network of Fig. 7.9, with

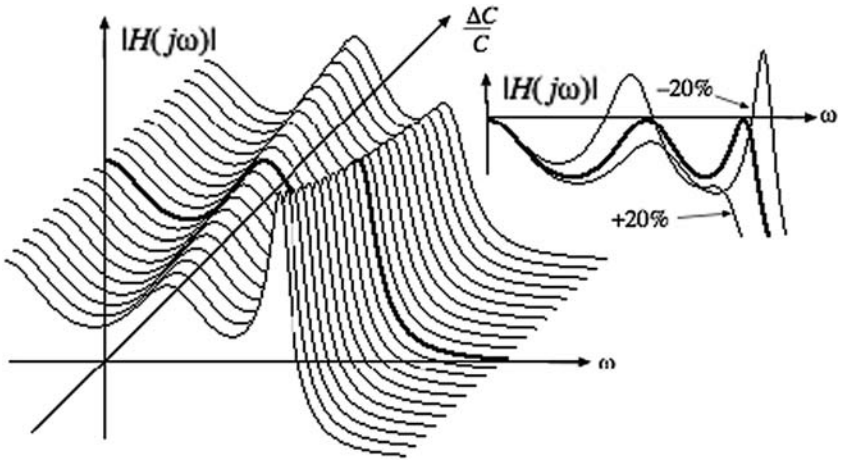


Fig. 7.11 Variation of the magnitude function due to ±20% variation in C_2

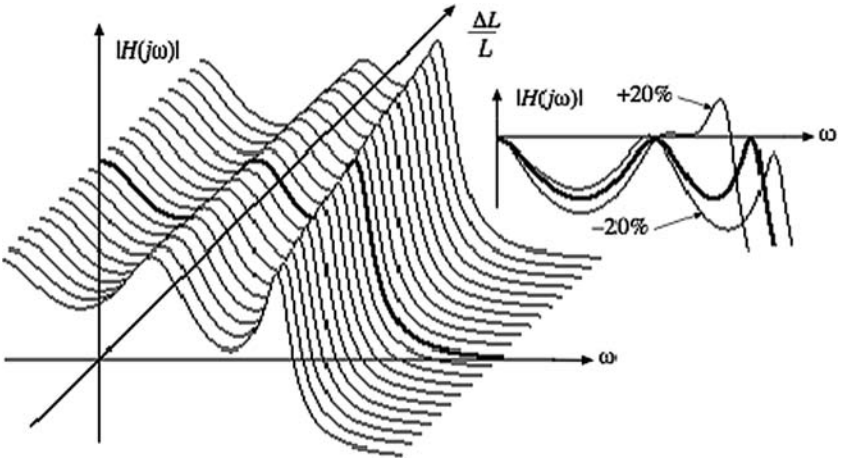


Fig. 7.12 Variation of the magnitude function due to ±20% variation in L_3

the filter in Example 7.1. The filter is a Chebyshev I filter with $A_{max} = 3$ dB. The element values varies ±20% around the nominal values.

7.3 Errors in the Elements in Doubly Terminated Filters

Figure 7.16 shows the typical deviation in the attenuation for a doubly resistively terminated LC filters due to errors in the reactive elements.

It can be shown that the deviation, shown in Fig. 7.16, in the passband attenuation for a doubly resistively terminated filter is [7]

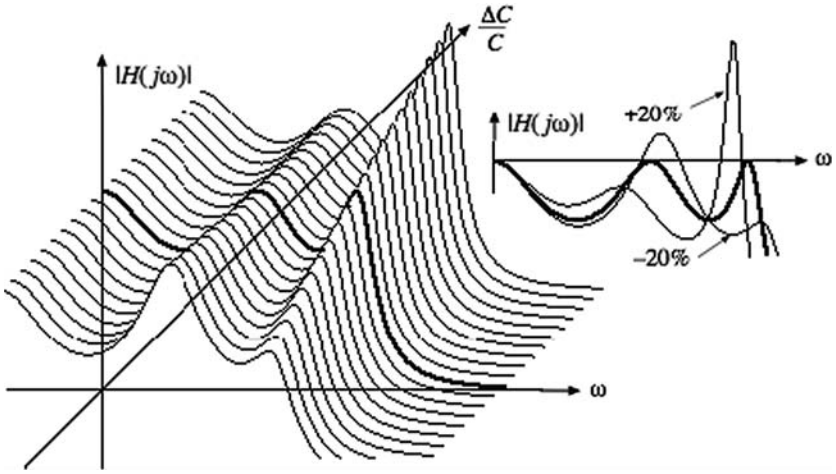


Fig. 7.13 Variation of the magnitude function due to ±20% variation in C_4

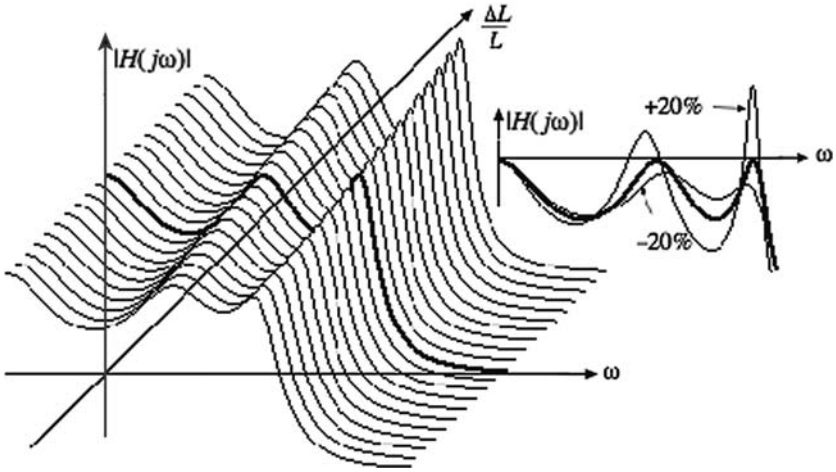


Fig. 7.14 Variation of the magnitude function due to ±20% variation in L_5

$$\Delta A(\omega) \leq 8.69\varepsilon \frac{|\rho(j\omega)|}{|H(j\omega|^2} \omega \tau_g(\omega) \tag{7.12}$$

where $\varepsilon = |\Delta L/L| = |\Delta C/C|$ represent the uniformly distributed errors in the inductors and the capacitors, i.e., $(1-\varepsilon)L \leq L \leq (1+\varepsilon)L$, etc. It can be shown that $\Delta A(\omega)$ is proportional to the electric and magnetic energy stored in the capacitors and inductors and that Eq. 7.11 also holds for commensurate transmission line filters. Note that Eq. 7.12 is not valid for singly terminated filters.

The deviation will, according to Eq. 7.12, be largest for frequencies where $\omega\tau_g(\omega)$ is largest, since the reflection function, $|\rho(j\omega)|$, is small and $|H(j\omega)| \approx 1$ in the passband. Hence, a doubly resistively terminated filter with 3 dB ripple in the

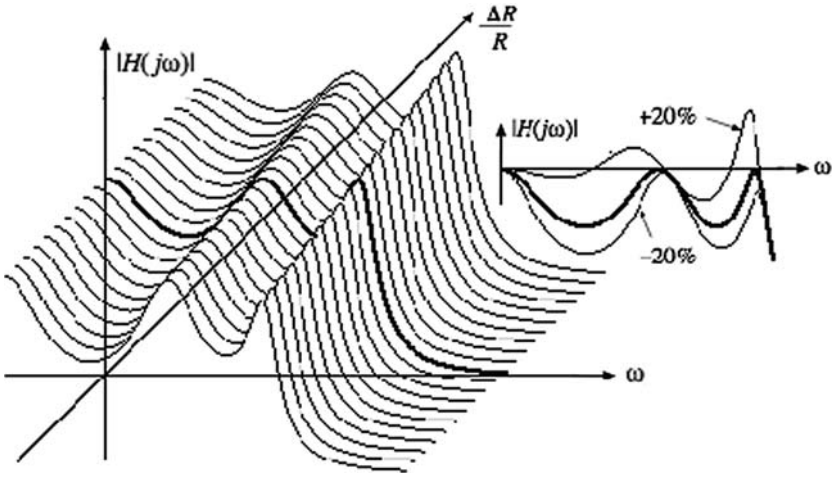
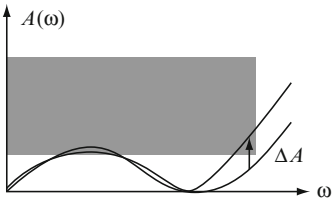


Fig. 7.15 Variation of the magnitude function due to ±20% variation in R_L

Fig. 7.16 Deviation of the attenuation due to errors in the reactive elements



passband is significantly more sensitive for element errors than a filter with smaller passband ripple, e.g., 0.01 dB. Moreover, the Q factors of the poles will be small if the passband ripple is small. Thus, it is often better to design a filter with a small ripple at the expense of a slightly higher filter order.

7.3.1 Errors in the Terminating Resistors

The sensitivities with respect to R_s and R_L are

$$S_{R_s}^{H(j\omega)} = -8.69\frac{\rho_1(j\omega)}{2} \tag{7.13}$$

and

$$S_{R_L}^{H(j\omega)} = -8.69\frac{\rho_2(j\omega)}{2} \tag{7.14}$$

Hence, the sensitivities are small in the passband, since $|\rho(j\omega)| << 1$, and equals zero for the frequencies at maximal power transfer. In addition, the errors will essentially appear as a small change in the gain of the filter and not affect the frequency selectivity.

7.3.2 *Effects of Lossy Elements*

The effect on the attenuation of lossy reactive elements can be estimated in terms of their Q factors where we assume that all inductor have the same Q factor and the same holds for the capacitors [6, 8].

$$\Delta A(\omega) \leq \frac{8.69}{2}\left(\frac{1}{Q_L}+\frac{1}{Q_C}\right)\omega\tau_g(\omega)+\frac{8.69}{2}\left(\frac{1}{Q_L}-\frac{1}{Q_C}\right)|\rho|_{\max} \qquad (7.15)$$

Also in this case it is favorable to select a small ripple in the passband. The deviation will be largest at the passband edge, i.e., where $\omega\tau_g(\omega)$ is largest.

Example 7.3. Consider a fifth-order Chebyshev I filter with A_{max}=0.5 dB, A_{min} = 42 dB, $\omega_c = 1$ rad/s, and $\omega_s = 2$ rad/s and we assume that the components errors are uniformly distributed with $\varepsilon = \pm 1\%$.

Figure 7.17 shows the deviations in the attenuation in the passband and stopband as well as the bound on the deviation according to Eq. 7.12. A significant number of filters do not meet the specification since the design margin is very small. In practice, a design margin should be allocated to the two bands as well as to the band edges. We will not discuss this related issue here.

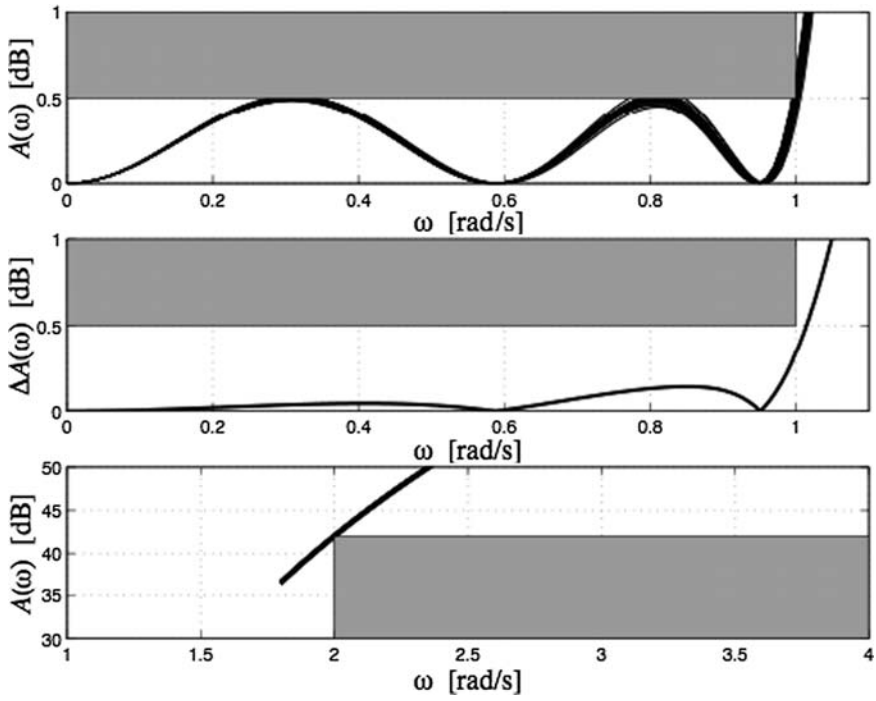


Fig. 7.17 Deviation in the passband attenuation (*top*), bound on the attenuation (*middle*), and deviation in the stopband attenuation (*bottom*)

Obviously, the deviation increases towards the passband edge while it is insignificant at low frequencies. Moreover, the sensitivity at the band edge is large and the cutoff frequency is sensitive to component errors.

7.4 LC Filters with Diminishing Ripple

Due to deviations in the attenuation caused by errors in the element values a part of the allowed ripple in the passband, A_{max} must be reserved to allow for errors in the component values. The filter must therefore be synthesized with a design margin, i.e., with a ripple, which is less than required by the application. According to Eq. 7.11, the deviation is smaller for low frequencies and increases towards the passband edge. In practice, however, in order to simplify the synthesis, the design margin is for the standard approximations distributed evenly in the passband even though the margin will not be exploited for lower frequencies.

In order to exploit the allowed passband ripple better we may let the reflection function $\rho(j\omega)$ of the synthesized filter decrease at the same rate as the other factors in $\Delta A(\omega)$ increase so that $A(\omega) + \Delta A(\omega) \leq A_{max}$. The ripple will decay towards the passband edge and the corresponding *LC* filter can be implemented with components with larger tolerances, i.e., the filter can be implemented with a lower overall cost. The group delay of a filter with diminishing ripples is slightly smaller than for the original filter. An additional advantage is that this will reduce the thermal noise as well [4].

Example 7.4. Consider a fifth-order Chebyshev I filter with $A_{max} = 0.5\,\text{dB}$, $\omega_c = 1\,\text{rad/s}$, and $\omega_s = 2\,\text{rad/s}$ and we assume that the components errors are uniformly distributed with $\varepsilon = \pm 3\%$.

Figure 7.18 shows the passband attenuation of the fifth-order regular filter in Example 7.2 and an approximation where the function $A(\omega) + \Delta A(\omega)$ is equiripple.

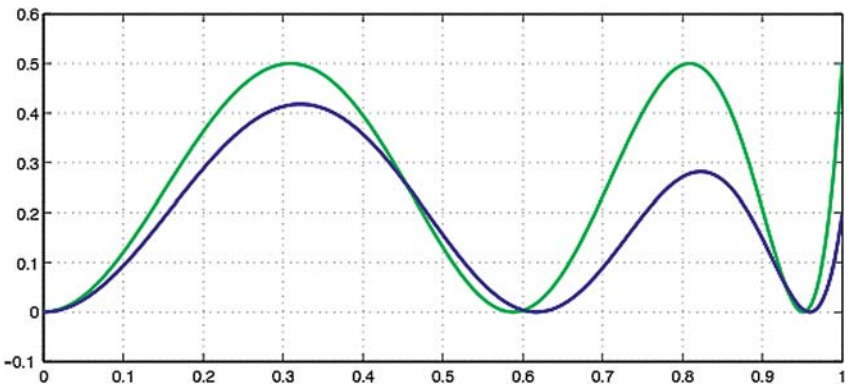


Fig. 7.18 Passband attenuation for fifth-order Chebyshev I filter with equiripple and diminishing ripple

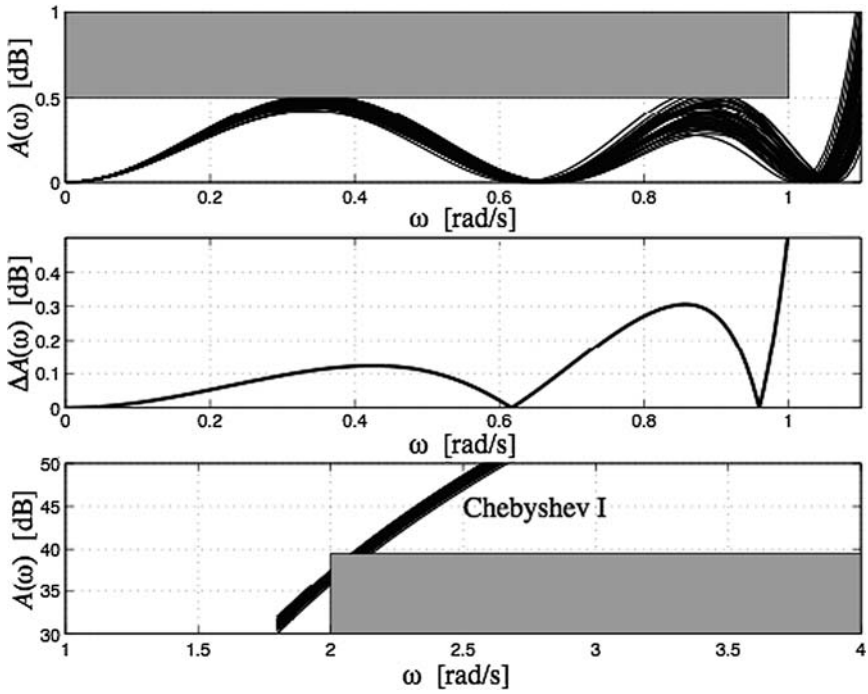


Fig. 7.19 Fifth-order Chebyshev I filter with equiripple and diminishing ripple

The attenuation has diminishing ripples in order to compensate for the increase in $\Delta A(\omega)$.

We get the following deviations in the passband and stopband attenuation: Figure 7.19. The acceptable components tolerances has been increased to $\varepsilon = \pm 3\%$ with about the same yield as in Example 7.2. In addition, the nominal passband edge has been increased. However, the stopband edge has to be increased for the diminishing ripple approximation. Alternatively the filter order may be increased.

Note that $\Delta A(\omega) = 0$ for the frequencies of maximum power transfer and that the nominal passband edge was increased in order to obtain a design margin at the cutoff edge where the sensitivity is largest.

7.5 Approximations with Small Group Delay

As seen above, it is favorable to select an approximation that has a small group delay. We therefore compare the group delay for the common standard approximations.

Example 7.5. Compare the Butterworth, Chebyshev I, Chebyshev II and Cauer standard approximations, which meet the same standard LP specification: $A_{max} = 0.00695\,\text{dB}(\rho = 4\%)$, $A_{min} = 45\,\text{dB}$, $\omega_c = 1\,\text{rad/s}$ and $\omega_s = 2\,\text{rad/s}$. Note that

A_{max} has been chosen very small in order to make the reflection function small, $|\rho(j\omega)|$, and make $|H(j\omega)|$ close to 1. Hence, the right hand side of Eq. 7.12 will be small and, in addition, the design margin will be large. Hence, we may use components with large tolerances ε at the expense of a slightly higher filter order.

We get the following filter orders with the four standard approximations.

Butterworth	$N_B = 12.12$	$\Rightarrow$	$N_B = 13$
Chebyshev I and Chebyshev II	$N_C = 6.904$	$\Rightarrow$	$N_C = 7$
Cauer	$N_{Ca} = 4.870$	$\Rightarrow$	$N_{Ca} = 5$

Note the large difference between the required orders for the standard approximations that meet the same requirement. The difference tends to increase if the transition band is reduced.

Figure 7.20 shows the attenuation for the different approximations. The allowed passband ripple is very small and we are only interested in that the requirement is met, and not in detailed variation inside the passband.

The attenuation in the transition band and the stopband varies between the different filters. The Chebyshev II filter has a more gradual transition between the passband and stopband compared with the Chebyshev I filter in spite of the fact that they have the same order. Note that the Cauer filter has a smaller transition band than required. The attenuation approaches in this case infinity for all of the filters.

Figure 7.21 shows the corresponding group delays. The difference in the group delay is large between the different approximations and the group delays maximum lies above the passband edge $\omega_c = 1$ rad/s. The Butterworth and Chebyshev I filters have larger group delay in the passband while the Chebyshev II and Cauer filters have considerably smaller group delay and the difference between the later two is relatively small.

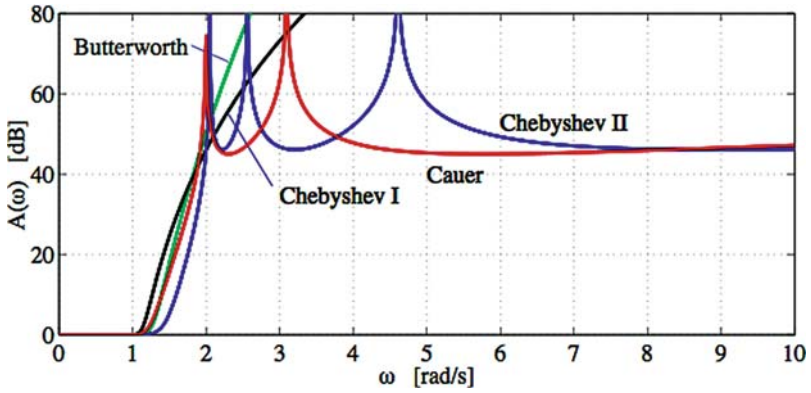


Fig. 7.20 Attenuation for the four approximations meeting the same specification

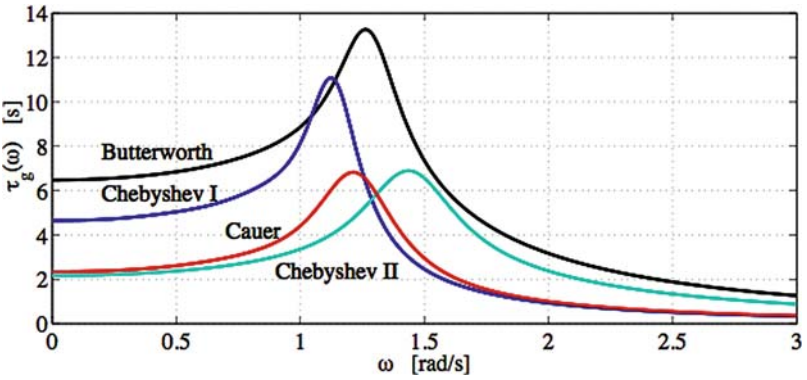


Fig. 7.21 Group delay for the four approximations

In literature, it is commonly stated that the Butterworth filter has the best group delay properties. This is obviously not correct and is based on an unfair comparison between the standard approximations of the same order. According to Fig. 7.21, Chebyshev II and Cauer filters have considerably better group delay properties.

The element sensitivity for an *LC* filter is proportional to the group delay.

The group delays at the passband edge and the passband variations are:

		$\tau_g(\omega_c) - \tau_g(0)$
Butterworth	8.87115	2.39415
Chebyshev I	8.15069	3.4990
Chebyshev II	3.36427	1.1934
Cauer	4.41338	2.0718

Notice that the Chebyshev II and Cauer approximations have the smallest variation and their smoothness are similar for the four approximations. The deviation due to losses and component errors, with $Q_L > 200$ and $Q_C > 50$, are:

	$\Delta A(\omega)$ according to Eq. 7.15 (dB)	$\Delta A(\omega)$ according to Eq. 7.12 (εdB)
Butterworth	0.4080	3.0880
Chebyshev I	0.2479	2.8380
Chebyshev II	0.1023	1.1713
Cauer	0.1342	1.5365

For example, by using a Chebyshev II or Cauer filter instead of a Butterworth filter the group delay can be reduced with a factor of 2.64 and 2, respectively, and the component tolerances can be increased with the same factor. An additional improvement can be obtained by using a diminishing ripple approximation that allocate

a larger design margin and reduces the sensitivity at the upper part of the passband. Moreover, it reduced the group delay and Q vales as well. Components with large tolerances are considerably cheaper than those with small tolerances. In addition, the number of components is fewer; 9 and 7 compared to 13 for Butterworth. It is therefore important to use an approximation with small group delay. Cauer is often the preferred approximation since the require order is significantly lower than for Chebyshev II and the group delay.

The Q factors for the four filters are shown below.

7.6 Design of Doubly Terminated *LC* Filters

Instead of using expensive components with low tolerances and large Q factor we can compensate for an increase in ε, i.e., using components with larger tolerances, using either or all of the following trade-off so the maximum of $A + \Delta A$ in the passband does not increase.

Notice, that the Cauer approximation yields the lowest Q values. The conclusion is that Cauer is the best approximation in most cases, i.e., when we have requirements on both the attenuation and group delay. In addition, the Cauer approximation yields *LC* filter with fewer components and with almost as low sensitivity to errors in the element values as Chebyshev II filters.

- Use a doubly resistively terminated reactance network that is designed for maximum power transfer, i.e., Eq. 7.11 is valid.
- Reduce $|\rho(j\omega)|$ by reducing the passband ripple, A_{max}, of the filter more than required by the application. However, this requires that the filter order to be increased. That is, we can use a few more, but cheaper components to reduce the overall cost of the implementation.
- Use an approximation that have low group delay, i.e., Chebyshev II and Cauer filters are preferred over Butterworth and Chebyshev I filters, see Fig. 7.21.
- Use an approximation with diminishing ripple.

Comparison of Q factors

Butterworth	Chebyshev I	Chebyshev II	Cauer
$N = 13$	$N = 7$	$N = 7$	$N = 5$
0.50000	0.50000	0.50000	0.50000
0.51494	0.68955	0.60835	0.83127
0.56468	1.33370	1.03162	3.12162
0.66799	4.34888	3.19218	
0.88018			
1.41002			
4.148115			

7.7 Conclusions

A singly resistively terminated ladder for which the principle of maximum power transfer is not valid has, thus, a high sensitivity in the passband. Such *LC* filters should therefore not be used since it would require more expensive components with smaller tolerances. A singly resistively terminated *LC* ladder network is much more sensitive for variations in the element values than a doubly resistively terminated *LC* ladder network. A doubly resistive terminated filter can be designed to have minimal element sensitivity in the passband. We presented an upper bound for such filters for the deviation in the passband. In addition we proposed a design strategy for low-sensitivity filters based on minimizing the group delay and ripple in the passband. Furthermore we compared the standard approximation with respect to the group delay and Q factors under the same magnitude requirements and found the best choice is either the Chebyshev II or Cauer approximation. By selecting Chebyshev II or Cauer approximation we get a reduction in the sensitivity by about a factor 2 and 2.6, respectively. Hence, components with 3 time higher tolerances may be used. Furthermore we showed that an addition improvement the sensitivity is obtained by using a diminishing ripple approximation.

References

1. L.T. Bruton, "*RC-Active Circuits, Theory and Design*", Prentice Hall, Englewood Cliffs, N.J., 1980
2. T. Deliyannis, Y. Sun, and J.K. Fidler, "*Continuous-Time Active Filter Design*", CRC Press, USA, 1999.
3. J. Gorski-Popiel, "*RC*-active synthesis using positive-immittance converters", *Electronics Letters*, Vol. 3, pp. 381–382, 1967.
4. G. Groenewold, "Noise and group delay in active filters", *IEEE Transactions on Circuits and Systems*, Part I, Vol. 54, No. 7, pp. 1471–1480, Jul. 2007.
5. H.J. Orchard, "Inductorless filters", *Electronics Letters*, Vol. 2, p. 224, Sept. 1966.
6. A.S. Sedra and P.O. Brackett,"*Filter Theory and Design: Active and Passive*", Pitman, London, 1978.
7. Wanhammar, "A bound on the passband deviation for symmetric and antimetric commensurate transmission line filters", http://www.es.isy.liu.se/publications/, 1991.
8. L. Wanhammar, "*Analog Filters Using MATLAB*", Springer, 2009.

Chapter 8
High-Performance Continuous-Time Filters with On-Chip Tuning

Jose Silva-Martinez and Aydın İ. Karşılayan

Abstract High performance continuous time filters based on operational transconductance amplifiers (OTAs) are discussed. Several OTA linearization techniques are reviewed, and a design example is provided with measurement results. To address the problem of inaccuracies in continuous time filters, two direct tuning techniques are presented with applications to ultra-wideband (UWB) receivers and bandpass sigma-delta modulators.

8.1 Introduction

Today's wireless receivers require very demanding high-performance analog filters that are typically used to block interferers and to provide anti-aliasing filtering before the subsequent analog-to-digital converters (ADCs). Furthermore, most of the high-performance continuous-time based oversampling ADCs targeted for broadband applications integrates the filter into the ADC loop. The filter still minimizes the power of the out-of-band interferers before the quantizer and provides noise shaping in filter's passband. ADC dynamic range is a strong function of filter's dynamic range [1, 2].

Typical filter implementations are based on active elements such as active-RC, MOSFET-C and G_m-C techniques [3–5]. For high frequency applications, however, G_m-C solutions are generally very competitive due to the efficient operation of wide-band operational transconductance amplifiers (OTA). However, OTAs usually present limited voltage-to-current conversion linearity. To overcome this problem, a wide variety of solutions has been reported during the last few years [4–15]. To further optimize OTA-C (or G_m-C) filter's performance, it is desirable to achieve a good balance between signal-to-noise ratio (SNR) and signal-to-distortion ratio (SDR).

J. Silva-Martinez and A.İ. Karşılayan (✉)
Department of Electrical and Computer Engineering, Texas A&M University College Station, Texas, USA
e-mail: karsilay@ece.tamu.edu

A.H.M. van Roermund et al. (eds.), *Analog Circuit Design: Smart Data Converters, Filters on Chip, Multimode Transmitters*, DOI 10.1007/978-90-481-3083-2_8,
© Springer Science+Business Media B.V. 2010

Therefore, the use of linearization schemes without sacrificing other important parameters such as noise level, power efficiency and frequency response becomes necessary.

OTA linearization techniques are discussed in Section 8.2, where a design example is also presented. In Sections 8.3 and 8.4, automatic tuning systems for two relevant applications are provided. Finally, some concluding remarks are given.

8.2 Linear Operational Transconductance Amplifiers (OTAs)

The differential pair shown in Fig. 8.1a is a basic cell that offers good rejection to both common mode input signals and power supply noise. Assuming that all transistors operate in saturation region, the large signal behavior of the single-ended output current can be described as follows:

$$i_d = \sum_{n=0}^{\infty} G_{MN\_(2n+1)} {v_{in}}^{2n+1} \tag{8.1}$$

In this expression, it is assumed that the even-order harmonic distortions are not prominent due to the fully-differential nature of the topology and the use of a common-mode feedback system. For moderate signal swing, the first two terms are the most significant ones; G_{MN_1} corresponds to the linear transconductance term and G_{MN_3} is the undesired third-order non-linear term. Values of G_{MN_1} and G_{MN_3} can be obtained using circuit simulations with complex transistor models such as BSIM.

A rough approximation based on the saturation square model and taking into account the mobility degradation effect due to the lateral and transversal electric fields [16] leads to,

$$G_{MN\_1} = \frac{1}{2} \frac{\sqrt{K_p \left({W_N}/{L_N} \right) I_T}}{1 + \frac{2}{\varepsilon_c} \sqrt{\frac{I_T}{W_N L_N K_p}}} \tag{8.2}$$

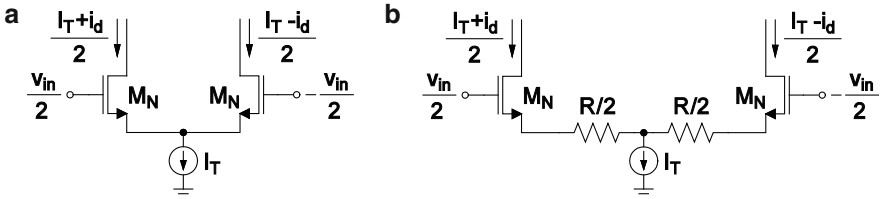


Fig. 8.1 **(a)** Conventional differential pair and **(b)** conventional source degeneration topology with the current source in the middle.

where K_p is a device parameter; W_N and L_N are the width and the length of the transistor M_N, respectively. ε_c is a fitting parameter determined by the critical electric field. The coefficient of the third order non-linearity is obtained as:

$$G_{MN\_3} = -\frac{G_{MN\_1}}{8\left(\frac{I_T L_N}{K_p \cdot W_N}\right)\left(1+\frac{2}{\varepsilon_c}\sqrt{\frac{I_T}{W_N L_N K_p}}\right)^3} \tag{8.3}$$

If a source degeneration resistance is added, as shown in Fig. 8.1b, the single-ended AC component of the output current approximately becomes [15]

$$i_{d1} \cong \left(1+\frac{G_{MN\_3}}{G_{MN\_1}(1+G_{MN\_1}R)^3}{v_{in}}^2\right)\left(\frac{G_{MN\_1}}{(1+G_{MN\_1}R)}\right)v_{in} \tag{8.4}$$

According to Eqs. 8.3 and 8.4, OTA's linearity can then be improved by increasing the source degeneration factor $G_{MN\_1}R$. However, this approach decreases the overall transconductance and usually results in higher noise levels and higher power consumption. Ideally the noise contribution of the tail current split equally in both branches and appears as common mode noise. Resistance and transistor mismatches present in the topology generate some differential output noise due to the tail current transistor, but usually noise increment is not relevant even if those mismatches are unrealistically large as 10%. Another good property of the circuit shown in Fig. 8.1b is that the non-linearities of the impedances lumped to the common-source node are not critical because that node remains almost unaffected by the differential signal variations. The drawback of the circuit shown in Fig. 8.1b is that the additional DC voltage drop across the resistors reduces the headroom for the input signal.

8.2.1 Advanced Linearization Techniques

It should be evident from Eq. 8.4 that increasing the transconductance of the main transistor leads to better linearity when source degeneration is employed. Techniques based on local feedback such as the one shown in Fig. 8.2a will improve OTA linearity provided that the voltage gain of the auxiliary amplifier is substantial

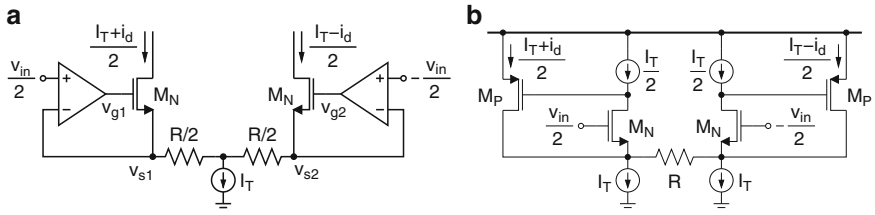


Fig. 8.2 Linearization techniques using **(a)** gate feedback **(b)** source feedback

in the frequency range of interest. In a first approximation, the transconductance gain of the combined amplifier-transistor circuit is given by $A_v(s)g_{mN}$. Although very wide linear range can be obtained with this topology, careful evaluation of circuit limitations is needed. First of all, notice that the DC level of v_{in} and $v_{s1,2}$ nodes is the same, while the DC level of $v_{g1,2}$ is V_{GS} volts higher. So, in the applications where the DC level of input and output are similar and the OTAs are cascaded, a level shifter may be needed [12]. This approach employs closed-loop regulation, therefore loop stability must be guaranteed. The poles at the nodes v_{si} and v_{gi} must be apart from each other to guarantee enough loop phase margin. Since $v_{s1,2}$ is a high frequency node, the frequency of the pole at $v_{g1,2}$ must be further reduced, limiting the high-frequency benefits of this technique. Evident downsides of this architecture are the significant overhead in power and silicon area. In addition, noise level usually increases due to the noise contribution of the auxiliary circuitry.

The architecture shown in Fig. 8.2b is based on the super-buffer topology [17]; it preserves the benefits of the regulated cascode topology and does not require any level shifter but a current mirror (not shown in the figure) to replicate the drain current of M_P transistors. The main concept behind this topology is to keep the drain current of M_N constant such that V_{GSN} is constant, hence the signal across R is identical to the input signal. The AC current generated by R is then collected by M_P and mirrored to have a very linear OTA. Linearity over -70 dB for signals of $0.5\,V_{pk-pk}$ can be achieved.

The third-order non-linearity can also be reduced by using special circuit techniques such as those reported in [6–14]. Before adopting any of the available linearization techniques, their drawbacks must be carefully evaluated. Some of the relevant issues are: (i) feedback techniques reduce the effective transconductance making it difficult to implement large transconductance values; (ii) silicon area and power consumption overhead can be significant when using excessive feedback; (iii) frequency response limitations due to additional poles present in signal path; phase errors are usually critical in highly selective filters; (iv) increased noise level; (v) compare the most relevant parameters: signal to noise and distortion ratios.

8.2.2 OTA Linearization Using Non-linear Elements

The non-linearity issue can also be alleviated using non-linear linearization methods. This approach has been used for some time in a number of applications such as translinear circuits. The origin of the non-linear behavior of the differential pair is the lack of current driving capability for large signals imposed by the tail current; this effect becomes evident when the static transfer characteristics of the differential pair are analyzed. This issue can be alleviated if the tail current tracks the amplitude of the input signal; thus more current is provided when the signal amplitude increases. A servo mechanism can do that job at low and medium frequencies but it may not be efficient for high speed applications due to the lack of speed of the servo-loop [16].

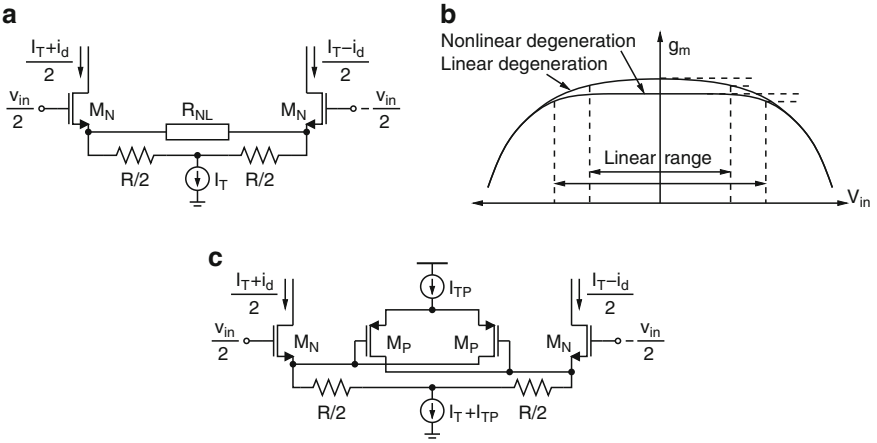


Fig. 8.3 Source degeneration using non-linear compensation **(a)** main concept, **(b)** typical large signal transconductance and **(c)** potential circuit implementation

Another possible solution uses a non-linear resistor as depicted in Fig. 8.3a. A negative non-linear resistance R_{NL} is connected in parallel with the linear source degeneration resistors. Although the formal analysis of the circuit is cumbersome, to get some insight on the operation of the circuit, the overall source degenerated resistance can be approximated as

$$R_{total} = \frac{1}{\frac{1}{R} - \frac{1}{R_{NL}}} = \frac{R}{1 - Rg_{M-NL}} \tag{8.5}$$

Usually $|R_{NL}| = |1/g_{M\text{-}NL}| >> R$ in order not to significantly degrade the overall transconductance. Since $R_{total} > R$, the overall OTA transconductance is slightly reduced due to the presence of the non-linear resistor R_{NL}. The overall transconductance can then be computed as

$$G_{m-total} = \frac{1}{\frac{1}{g_{M1}} + R_{total}} \cong \frac{g_{M1}}{1 + \frac{Rg_{M1}}{1 - Rg_{M-NL}}} = \frac{(1 - Rg_{M-NL})\, g_{M1}}{1 + R\,(g_{M1} - g_{M-NL})} \tag{8.6}$$

where both g_{M1} and $g_{M\text{-}NL}$ are non-linear functions. The overall transconductance reduces (usually less than 10%) as depicted in Fig. 8.3b. If the non-linear terms of both g_{M1} and $g_{M\text{-}NL}$ are matched, then the denominator becomes linear. For large input signals both, g_{M1} and $g_{M\text{-}NL}$, decrease leading to a partial non-linearity cancellation in the numerator too, resulting in a more linear voltage to current conversion. A simple servo-mechanism drastically reduces the sensitivity of the compensated OTA under large process-voltage-temperature variations [15]. Experimental results have shown that linearity improves over 10 dB. Noise overhead is small since the noise of the auxiliary tail current is common-mode and then rejected by the dif-

ferential nature of the architecture. Noise due to M_p is very small because the transconductance of the compensating circuit is smaller than one tenth of that of the main stage. Power consumption increases by less than 10%.

This technique can be easily implemented using an auxiliary differential pair (ADP). The inherently non-linear characteristics of the ADP are used to compensate the non-linearities of the main device attaching it in an anti-parallel configuration as depicted in Fig. 8.3c. The overall effect of the ADP is to generate a non-linear resistor that modulates the overall source degeneration resistance; thus the source degeneration factor varies with the amplitude of the input signal such that the variations of the overall transconductance are minimized.

A major issue in this topology is the design of the ADP; a design strategy can be found in [15]. For a $0.7\,V_{pk-pk}$ input, the conventional source degenerated differential pair without ADP shows a third-order intermodulation distortion (IM3) of -53 dB while with the ADP activated, only one sixteenth additional power ($I_T = 16I_{TP}$), the IM3 is reduced down to -67 dB for frequencies up to 60 MHz. The size of the ADP requires 16% additional area and the overall small-signal transconductance decreases by less than 10%. To evaluate the effects of process parameter variations on the linearity of the compensated OTA, Monte Carlo simulations were carried out using random process variations around the corners for three values of R. Main technological parameter variations ($\pm 10\%$) include oxide thickness, threshold voltage, width and length etching effects. Figure 8.4 shows a scatter plot relating the third intermodulation component of the differential pair with ADP ($IM3_{ADP}$) activated against the linearity of the differential pair without ADP ($IM3_n$). For most of the cases a linearity improvement of at least 10 dB can be guaranteed over all technology corners.

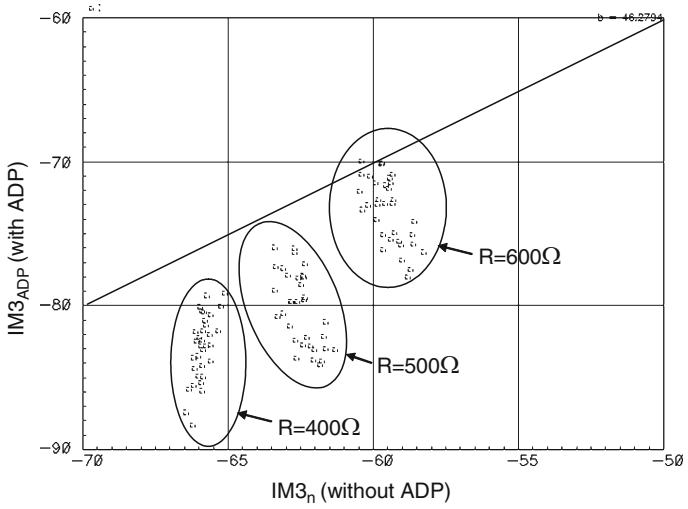


Fig. 8.4 Scatter plot of Monte Carlo simulation of IM3 with and without ADP including process parameter variations. Nominal $R = 500\,\Omega$

8.2.3 Design Example: A 30-MHz Elliptic Filter

A 30 MHz fifth-order low-pass elliptic filter with 30 dB of stop-band attenuation suitable for power line communications is used as test bed. A robust solution can be achieved if implemented using the ladder structure shown in Fig. 8.5. Four OTAs are needed for the realization of each floating inductor. The minimum capacitance (2.3 pF) was selected based on thermal noise and matching considerations. The filter uses the same OTA for all resonators to improve the matching of the time constants; impedances were scaled to match the conductance of filter's terminations with OTA's transconductance. A self-bias circuit matches the OTA transconductance to the resistance value to make the topology immune to PVT variations; polysilicon resistors are used as filter terminations. The OTAs are implemented using the architecture shown in Fig. 8.3c, terminated by a couple of current sources; the absence of a high impedance output stage in the OTAs allows us to save power and reduce noise at the expense of limited DC gain. The finite OTA's output resistance introduces some losses in the resonators, limiting the attenuation at the location of the high-Q zeros. However, both roll-off and stop-band attenuation remained within the specifications. A common-mode feedback is necessary to fix the OTA output common-mode level and to reject power supply noise. OTAs are not tuned to avoid linearity degradation; instead banks of capacitors are used for filter's frequency tuning. Design details can be found in [15].

The circuit was fabricated in the TSMC 0.35 μm CMOS process. Figure 8.6 shows the micrograph of the chip; total active area is 1.4 mm^2. The supply voltage is 3.3 V and total power consumption of the filter is 85 mW. Figure 8.7a shows all the low-pass responses that can be programmed using the capacitor banks. The frequency control results in sixteen cut-off frequencies from 20 to 33 MHz. The roll-off of the passband to stop band is approximately 35 dB per octave. Figure 8.7b shows the in-band third intermodulation distortion (IM3) test for a signal composed by two tones of 0.5 V_{pk-pk} each around 20.5 MHz. Two shifted graphics are overlapped for comparison; the tones indicated with arrows correspond to the test with the ADP

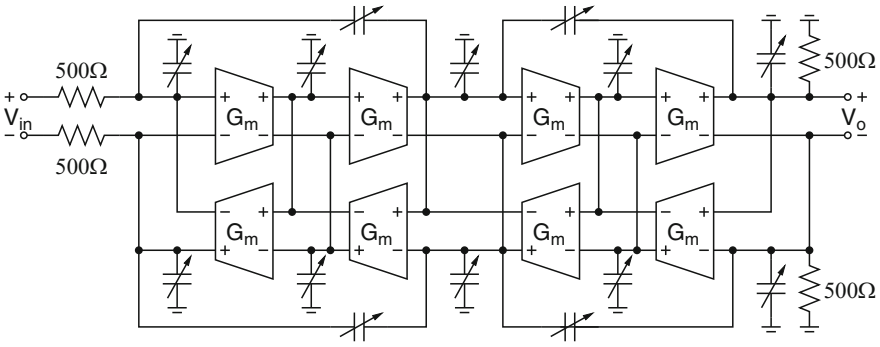


Fig. 8.5 Schematic of the resistively terminated fifth-order elliptic ladder filter

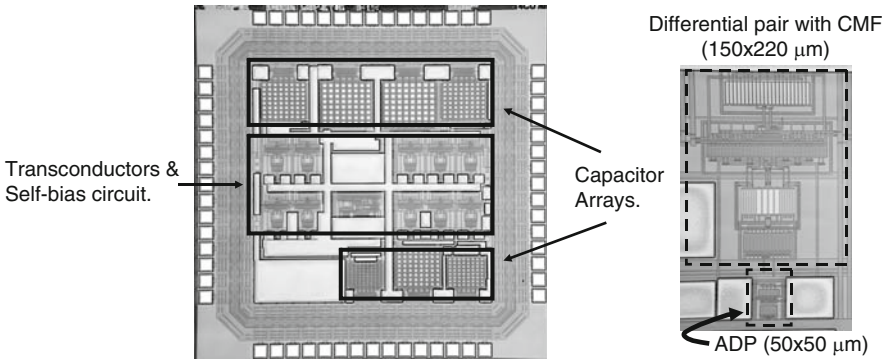


Fig. 8.6 Chip micrograph. Area of the OTA $= 150 \times 220\,\mu m^2$ and ADP $= 50 \times 50\,\mu m^2$

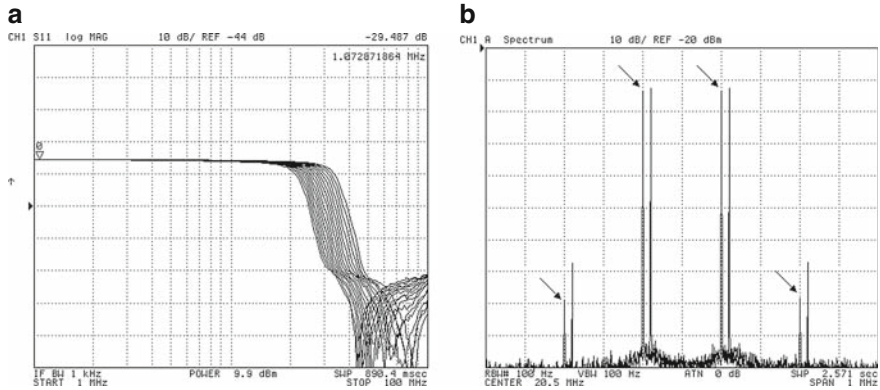

a
CH1 S11 log MAG 10 dB/ REF -44 dB -29.487 dB
1.872871868 MHz
IF BW 1 kHz POWER 9.9 dBm SWP 890.4 msec
START 1 MHz STOP 100 MHz
b
CH1 A Spectrum 10 dB/ REF -20 dBm
RBW 100 Hz VBW 100 Hz ATN 0 dB SWP 2.571 sec
CENTER 20.5 MHz SPAN 1 MHz

Fig. 8.7 **(a)** Programmable cut-off frequencies of the filter, **(b)** IM3 test; the tones with the arrow are for the filter with ADP activated

activated; as expected, IM3 improves by more than 10 dB when the compensation network is activated. The power consumption is almost the same for both cases. Signal to distortion ratio increases by 10 dB while signal to noise ratio and power consumption are roughly the same. Near and far stop band harmonic distortions folded back into baseband were found to be less than −76 dB. The experimental results are summarized in Table 8.1.

8.3 Broadband Tuning for Interference Suppression in UWB Receivers

Ultra-wideband (UWB) receivers are susceptible to in-band narrowband interference (NBI) due to their low power and wide bandwidth [18, 19]. After down-conversion of the RF signal in the receiver, interference could occupy any frequency

Table 8.1 Summary of filter results

Power consumption (filter only)	75 mW
Integrated input referred noise	115 μV
IM3 @ 1 V_{pk-pk}, 20 MHz (in-band)	−75 dB
IM3 @ 1 V_{pk-pk}, 30 MHz (in-band)	−60 dB
IM3 @ 1 V_{pk-pk}, 40 MHz, 50 MHz. (near out-band)	−76 dB
IM3 @ 1 V_{pk-pk}, 60 MHz, 90 MHz. (far out-band)	−90 dB
SNR @ 1 V_{pk-pk} input	67 dB
SFDR @ 30 MHz	65 dB
PSRR+; PSRR−	>30 dB

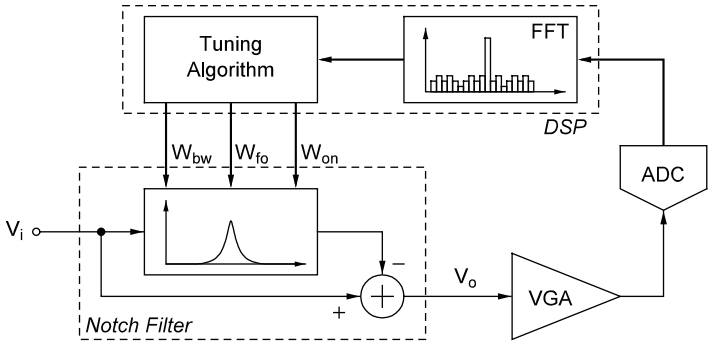


Fig. 8.8 Notch filter utilizing the DSP's FFT block for interference detection and center frequency tuning of the notch filter

within the UWB baseband from a few MHz up to 264 MHz, which can be reduced using a programmable analog notch filter before the signal is quantized. The notch filter's center frequency (ω_0) should be adjustable throughout this range, and a method should exist for adaptively changing ω_0 to match the interference frequency. Detecting the existence and location of the interference is a challenging task in the analog domain, but becomes relatively straightforward with the use of an FFT. Multiband-OFDM based UWB systems already use an FFT processor to decode the UWB data, so interference detection can be done with minimal additional overhead. If digital interference detection via FFT is used, for simplicity it follows that the notch filter's center frequency should have digital controls.

The block diagram of the filter with the center frequency control included is shown in Fig. 8.8, where the notch filter architecture is based on feedforward subtraction of the bandpass filtered signal. The digital control signals W_{fo}, W_{bw}, and W_{on} are generated by the DSP when interference is detected.

Figure 8.9 shows the digitally controlled notch filter utilizing banks of operational transconductance amplifiers (OTA). The bandpass filter is realized using an OTA-C biquadratic section, where cross-coupled OTA implements the subtraction to obtain the notch response. The digital controls W_{fo}, W_{bw}, and W_{on} represent the multiplication factors for the corresponding OTAs.

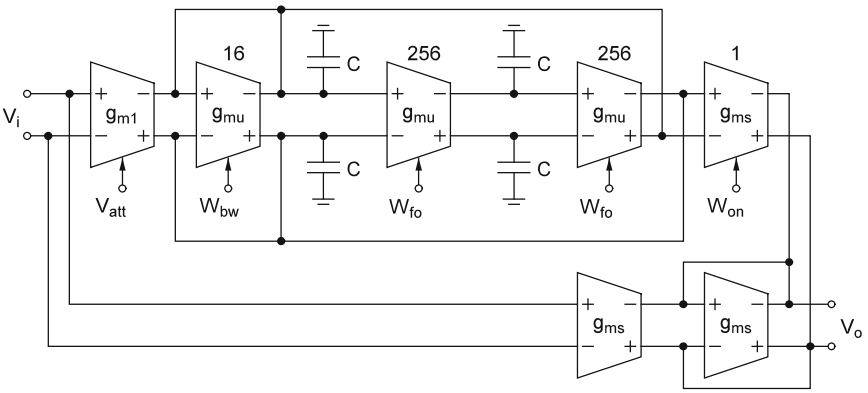


Fig. 8.9 Notch filter schematic with digital controls W_{fo}, W_{bw}, and W_{on} and analog control V_{att}. The numbers above the OTAs represent the quantity of unit OTAs in the digitally controlled bank of OTAs

Transfer function of the notch filter in Fig. 8.2 can be obtained as,

$$H_N = \frac{s^2 + \frac{W_{bw}g_{mu}+2W_{fo}g_{ou}-g_{m1}}{C}s + \frac{W_{fo}^2 g_{mu}^2}{C^2}}{s^2 + \frac{W_{bw}g_{mu}+2W_{fo}g_{ou}}{C}s + \frac{W_{fo}^2 g_{mu}^2}{C^2}} \tag{8.7}$$

Based on Eq. 8.7, center frequency (ω_0), -3 dB bandwidth (ω_{bw}), and the attenuation at ω_0 (α_0) of the notch filter can be expressed as

$$\omega_0 = W_{fo}\frac{g_{mu}}{C} \tag{8.8}$$

$$\omega_{bw} = \frac{W_{bw}g_{mu} + 2W_{fo}g_{ou}}{C} \tag{8.9}$$

$$\alpha_0 = \frac{W_{bw}g_{mu} + 2W_{fo}g_{ou}}{W_{bw}g_{mu} + 2W_{fo}g_{ou} - g_{m1}} \tag{8.10}$$

From Eq. 8.2, it is clear that W_{fo} can be used to discretely control the center frequency of the filter with the step size g_{mu}/C. To ensure an attenuation of at least α_{min}, g_{mu}/C should be limited to

$$\frac{g_{mu}}{C} \leq \frac{\omega_{bw}}{\alpha_{\min}} \tag{8.11}$$

The number of required discrete frequency steps, N, can then be determined by

$$N \geq \frac{\omega_{\max}}{g_{mu}/C} \tag{8.12}$$

where ω_{max} is the maximum interference frequency. For this application, N is chosen as 256 with $\omega_{bw} = 2\pi 20$ Mrad/s, $\alpha_0 \geq 20$ dB, and $g_{mu}/C = 2\pi 2$ Mrad/s.

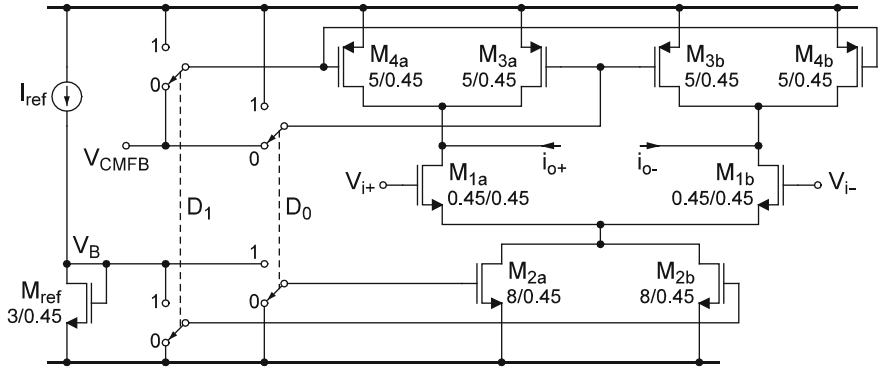


Fig. 8.10 Switchable unit OTA used in the notch filter

As the center frequency is tuned by varying W_{fo}, it is desirable to keep the bandwidth (ω_{bw}) constant. However, according to Eq. 8.9, ω_{bw} depends on the output conductance g_{ou} and W_{fo}. To keep ω_{bw} constant, W_{bw} should also be adjusted when W_{fo} changes. However, a maximum value of 16 is sufficient for W_{bw} due to relatively small value of g_{uo} compared to g_{mu}. The resulting W_{bw} can be calculated as

$$W_{bw} = 16 - \frac{W_{fo}}{16} \tag{8.13}$$

Schematic of the switchable unit OTA used in the notch filter is shown in Fig. 8.10. In this circuit, V_B is a DC bias generated from the current reference formed by I_{ref} and M_{ref}, and V_{CMFB} is generated from a common mode feedback circuit. The digital inputs D_1 and D_2 control the switches as indicated. When $D_1D_0 = 00$, the gates of the biasing transistors M_{2a} and M_{2b} are switched to ground, and the gates of the common mode feedback transistors, $M_{3a,b}$ and $M_{4a,b}$ are switched to V_{DD}. This effectively turns off the OTA by driving no bias current through the driver transistors, M_{1a} and M_{1b}. When $D_1D_0 = 01$, the gate voltage of M_{2a} becomes V_B, forming a current mirror with M_{ref} and the OTA is biased with a current equal to I_{ref}. The gates of M_{3a} and M_{3b} are tied to V_{CMFB} supplying the common mode feedback current to the OTA. Finally, if $D_1D_0 = 11$, transistors M_{2b}, M_{4a}, and M_{4b} turn on, and the OTA is biased with a current equal to $2I_{ref}$. Because there are three unique states for this OTA, one of these building blocks represent two unit OTAs, g_{mu} from Fig. 8.9. Note that $D_1D_0 = 01$ is the same as $D_1D_0 = 10$ because D_1 and D_0 control transistors with equal sizes.

8.3.1 Analog LMS Control for Maximizing Attenuation

In the presence of process and temperature variations and mismatch, the design of a bandpass filter with precisely unity passband gain becomes a challenging task.

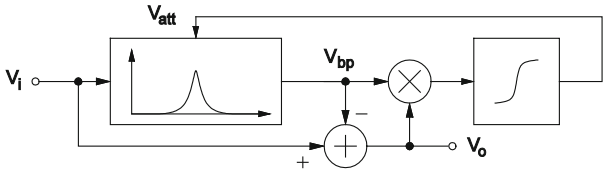


Fig. 8.11 LMS gain control for the notch filter

The fact that the filter's center frequency may change from a few MHz to 264 MHz further increases the challenge because unity passband gain needs to be ensured for all filter settings. It would be beneficial to use the available DSP for tuning the passband gain of the filter. However, in order to do so, the center frequency of the filter would already need to be tuned to the NBI. Unfortunately, proper operation of the center frequency tuning is conditioned on the passband gain already being unity. To address this issue, an analog Least Means Squared (LMS) tuning technique [20] can be applied to tune the passband gain of the filter. The additional overhead is one multiplier and one integrator to the existing filter architecture of Fig. 8.8. The block level schematic of the notch filter with LMS feedback is shown in Fig. 8.11.

Assuming the bandpass filter is properly tuned with center frequency equal to the interference frequency, and noting that $V_o = V_i - V_{bp}$ from Fig. 8.11, the analog control voltage V_{att} can be given as

$$V_{att} = M \int V_{bp}(V_i - V_{bp}) \tag{8.14}$$

V_{att} controls the passband gain of the filter, and accordingly, ω_0. If the gain of the bandpass filter is less than unity, $|V_{bp}|$ will be less than $|V_i|$, and the interference component of V_o will be in phase with V_{bp}. Integration of their positive product will result in increasing V_{att}. On the other hand, if the gain of the bandpass filter is greater than unity, the interference component of V_o will be 180° out of phase with V_{bp}. Their product will thus be negative, and integration will result in decreasing V_{att}, and in turn, the gain of the bandpass filter. Only when the gain of the bandpass filter is precisely unity will the product of V_o and V_{bp} equal zero, and thus no change in V_{att} will occur. One assumption for this circuit to work well is that the power of the interference is much greater than the UWB signal. This is a valid assumption because if the interference power is not much greater than the UWB signal power, the filter would not be needed.

Multipliers based on transconductance cells are an attractive option for implementing the multiplier because the output is a current. The integrator cell in Fig. 8.4 can thus be implemented by simply adding a capacitor to the output of the mixer. The schematic of the multiplier [21] is provided in Fig. 8.12. V_P and V_{CM} are DC biasing voltages generated from a current reference whereas V_{CMFBn} is generated from a common-mode feedback circuit. W_{on} is the same digital control that was used in Fig. 8.8. During symbol periods where there is no interference, $W_{on} = 0$, and V_{att} will be stored on the integrating capacitor. When the symbol period that contains the

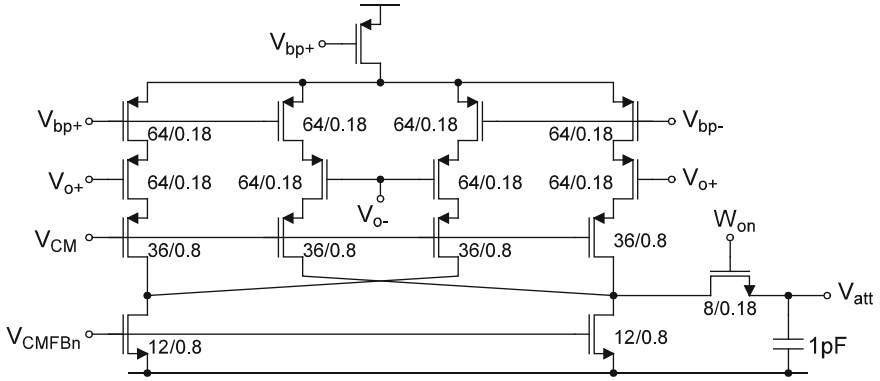


Fig. 8.12 Schematic of the multiplier/integrator

interference becomes active again, V_{att} will already be near the correct value, and $W_{on} = 1$ to resume the LMS convergence. V_{bp}, V_o, and V_{att} are connected to their respective nodes in Fig. 8.9.

8.3.2 Interference Detection and Center Frequency Tuning

The tuning algorithm block in Fig. 8.8 is completely in software and can thus be tailored to the designer's preference. The algorithm is responsible for both interference detection and filter center frequency tuning. Interference can be detected at the FFT output by comparing the peak FFT bin amplitude to the average. If this difference is above some threshold, this FFT bin is considered interference, and the notch filter is turned on. For tuning, various algorithms could be used. A brute force linear search is the simplest method; however, the time required for convergence could be quite large for high frequency interference. To speed up convergence, an initial guess can be made at the correct setting of W_{fo}. Since the guess could actually set the center frequency either too high or too low, the search should now progress in an outward fashion. In this algorithm, the control word, W_{fo}, is initially set to $W_{fo} = f_{int}/\rho$, where $\rho = g_{mu}/C$ is the frequency step. The value of ρ may be adaptively learned, and is initially set to the expected value from simulations. After the filter has settled, the FFT is taken again. If the interference was sufficiently attenuated, the filter is considered tuned, and the algorithm is complete. In all likelihood, due to process and temperature variations, the interference will still exist in the FFT output. If this is the case, W_{fo} is decreased by one, and the FFT is taken again. If the interference still exists, W_{fo} is increased by 2, and pending an incorrect control, will be decreased by 3 such that on the k^{th} attempt, the control word will be:

$$W_{fo}[k] = W_{fo}[k-1] + (-1)^{k-1}(k-1) \tag{8.15}$$

This outward search will continue until the interference is sufficiently attenuated in the output of the FFT. Once the filter has been properly adjusted, a new value can be computed for the step size, according to $\rho = W_{fo}/f_{int}$. This value will be used next time interference is detected. By adaptively changing ρ, the time required for convergence will be reduced for subsequent interference because the control of the filter is effectively learned by the algorithm. The proposed scheme is best utilized on static interference. In cases of frequency hopping NBI that frequently changes carriers, the tuning time may be too high, and alternative methods must be used. If it were possible to detect both the interference and the notch in the FFT output, faster tuning algorithms, such as binary searches, could be used. However, since the UWB input power could be much less than the interference, the notch response could be buried under the quantization noise of the ADC.

8.3.3 Experimental Results

The complete notch filter including center frequency adjustability and LMS tuning for maximum attenuation was fabricated in TSMC 0.18 μ technology. A micrograph of the filter is displayed in Fig. 8.13a. The shown micrograph consists of three notch filters, each having area $750 \times 500\,\mu m = 0.375\,mm^2$.

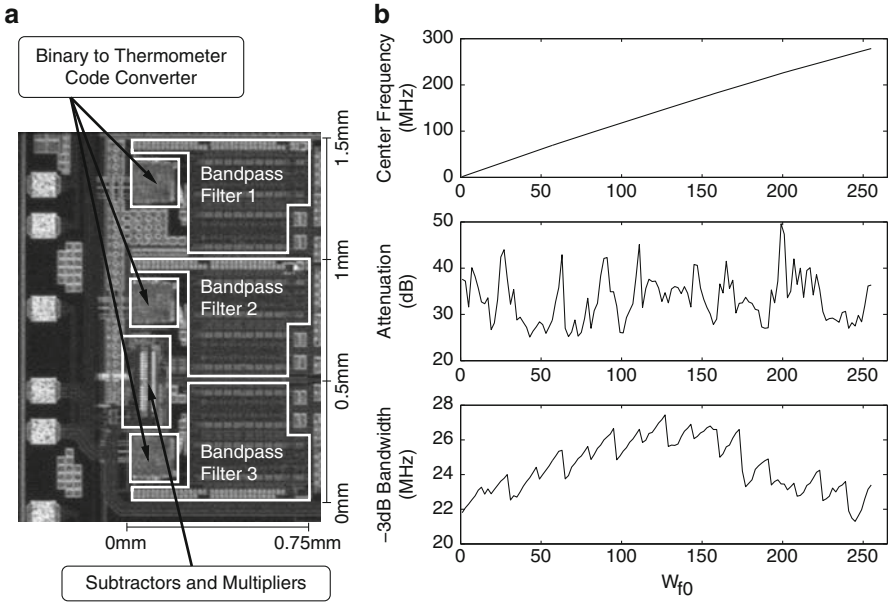


Fig. 8.13 **(a)** Chip micrograph and **(b)** measured filter characteristics versus the frequency control word, W_{f0}

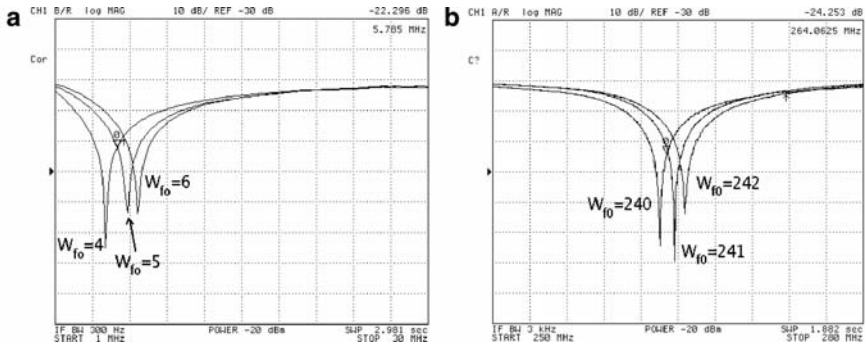


Fig. 8.14 Measured magnitude response for three adjacent high frequency W_{f0} settings of the notch filter **(a)** low frequency and **(b)** high frequency

Figure 8.13b shows the filter's characteristics versus the digital frequency control word, W_{fo}. The center frequency varied linearly with W_{fo} from 1.63 to 278.6 MHz. The worst case attenuation was 25 dB, and the best case was nearly 50 dB. The bandwidth was relatively constant, varying from 22 MHz to 27 MHz. Figure 8.14a displays the magnitude response of the notch filter for three adjacent low frequency W_{fo} settings. Figure 8.14b displays the magnitude response of the notch filter for the three adjacent high frequency W_{f0} settings. In both cases the frequency notches overlap below −20 dB, thus it is possible to attenuate any frequency within this continuous band by at least 20 dB. MB-OFDM is a frequency hopping system, and thus RF interference only appears in the baseband periodically. To emulate the intermittent nature of interference in MB-OFDM, the signal was applied to the input of the filter only for one interval. Figure 8.15a displays the resulting settling behavior at the filter's output when the filter is left on during all symbols and is subject to a practical interference situation. As expected, residual settling is observed in the filter output once the interference disappears from the filter input. By turning off the feedforward path of the notch filter, this residual settling can be avoided. Recall that the feedforward path can be turned off with the digital control W_{on} from Fig. 8.9. The measured settling behavior when the filter is turned on only during the symbol that contains interference is provided in Fig. 8.15b. As expected, there is no settling seen in the adjacent symbol.

8.4 Calibration of the Noise Transfer Function in a BP ΣΔ Modulator

In this section, a calibration technique for the Noise Transfer Function (NTF) optimization of Continuous-Time Bandpass Sigma Delta (CT BP) modulators is discussed. The technique employs test tones applied at the input of the quantizer to evaluate the noise transfer function of the Analog-to-Digital Converter (ADC)

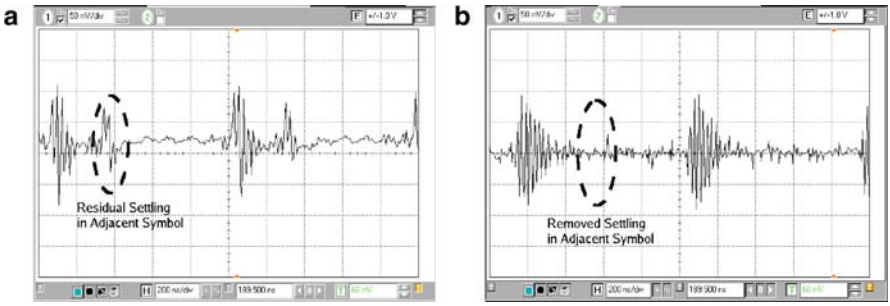


Fig. 8.15 Settling behavior of the notch filter's output, V_o, for the periodic interference input. **(a)** The notch filter is left on during all symbols. Residual settling is seen in the symbol period adjacent to the interference symbol. **(b)** The notch filter is turned off during the symbol periods that do not contain the interference. No residual settling is observed in the adjacent symbol

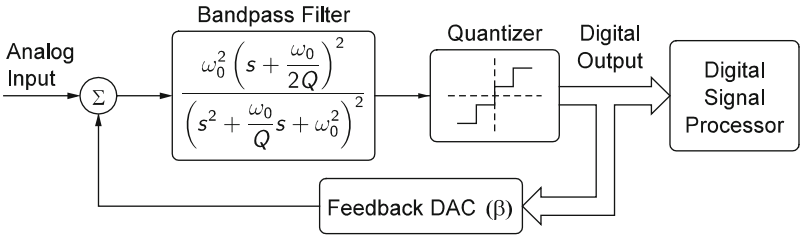


Fig. 8.16 Simplified block diagram of a continuous-time fourth order bandpass sigma-delta modulator

using the capabilities of the Digital Signal Processing (DSP) platform usually available in mixed-mode systems. Once the ADC output bit stream is captured, necessary information to generate the control signals to tune the ADC parameters for best Signal-to-Quantization Noise Ratio (SQNR) performance is extracted via an LMS software-based algorithm. The NTF notch frequency can be tuned using this methodology. The global calibration approach can be used during the system start-up and the idle system time. Figure 8.16 shows a simplified block diagram of a typical continuous-time BP Modulator using a fourth order BP filter and a multi-bit quantizer.

It can be shown that the quantization Noise Transfer Function (NTF) can be approximated as

$$NTF(s) = \frac{\left(s^2 + \frac{\omega_0}{Q}s + \omega_0{}^2\right)^2}{\left(s^2 + \frac{\omega_0}{Q}s + \omega_0{}^2\right)^2 + (\omega_0{}^2\beta)\left(s + \frac{\omega_0}{2Q}\right)^2} \tag{8.16}$$

where β is an expression dependent on the feedback DAC coefficient in the modulator; ω_0 and Q are filter's center frequency and pole's finite quality factor, respectively. Assuming ideal components the Q factor can be very large, and the

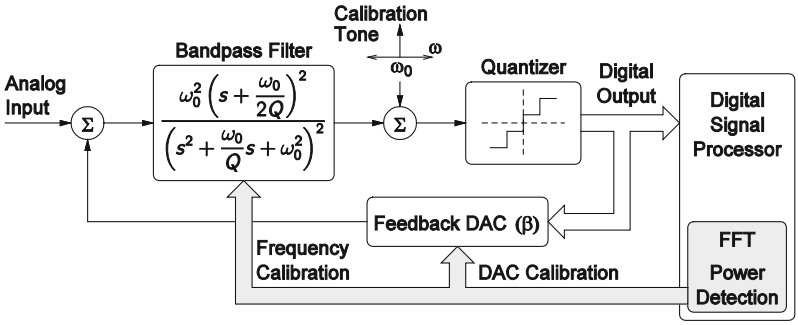


Fig. 8.17 BP-ΣΔ modulator with frequency and DAC calibration

magnitude of the NTF evaluated at the resonant frequency $\omega = \omega_0$ is zero, which leads to an excellent SNR performance around the resonant frequency. However, ω_0 in continuous-time filters typically changes by $\pm 25\%$ over process-voltage-temperature (PVT) variations. Also, a finite gain of <30 dB in single stage amplifiers and parasitic poles in high-gain amplifiers will limit the Q factor, which reduces the ADC's SNR. In addition, the excess loop delay between the quantizer sampling time and the time when a change in the output bit is seen at the feedback point in the filter that will cause SNR degradation and stability issues. A robust loop tuning approach should rely on a software-based platform instead of inaccurate analog circuitry. A system level implementation of the digital based tuning scheme for the ADC is shown in Fig. 8.17. In addition to an out-of band analog input, a test tone at the desired center frequency ω_0 is applied at the input of the quantizer to emulate a systematic and testable in-band quantization noise [22]. The quantizer output digital bit stream is then processed by the digital signal processor (DSP), and the power of the test tone is then measured in the digital domain using the Fast Fourier transform (FFT).

The estimated power of the test tone is used in an adaptive Least Mean Square (LMS) algorithm that controls several parameters with the aim of minimizing the power of the measured test tone and thus maximizing the rejection to quantization noise. The LMS algorithm generates the digital control signals to tune loop's notch frequency by controlling a bank of capacitors used for the realization of the bandpass filter. Once the notch frequency of the NTF is set at the desired frequency, the DAC coefficients and excess loop delay are then adjusted with the same aim: power minimization of the test tone to reach the best possible signal to quantization noise ratio. The calibration algorithm consists of the following steps: (i) inject an input signal such that the loop operates properly, frequency and power of this signal is not relevant for the operation of the calibration algorithm; (ii) Inject test tones at the desired frequencies at the quantizer input; (iii) Find the frequency component of the test tone frequency; (iv) By means of a LMS algorithm a digital control tuning signal is computed based on the difference between the stored and the new estimated power value of the detected test tones; (v) the parameters that control ω_0 are tuned first; (vi) iterate between (iii)–(v) until the power of the measured tone is minimized;

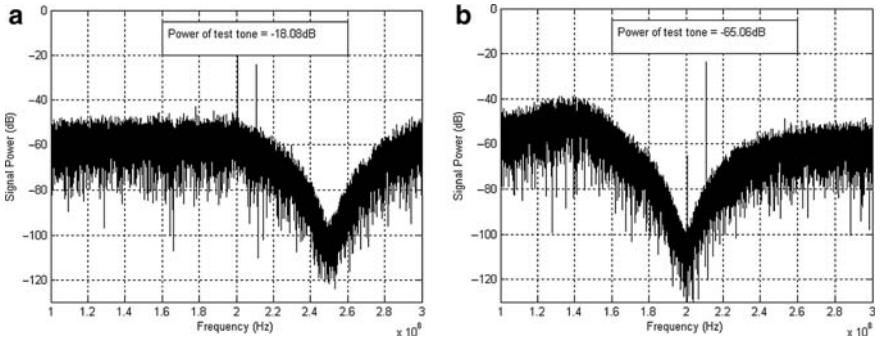


Fig. 8.18 **(a)** Output spectrum for the un-calibrated loop: PVT variations are over 25%; calibration tone is placed at 200 MHz. **(b)** Spectrum after calibration: power of calibration tone centered at 200 MHz is −65 dB

(vii) Once the frequency of the NTF notch is tuned, the algorithm tunes the DAC coefficients and a programmable delay element, if required, until the power of the detected test tone is minimized.

Figure 8.18a shows the response of the un-calibrated $f_s/4$ fourth order 200 MHz ADC to two tones. The first one is applied at the input of the ADC at 210 MHz; a calibration tone is applied at the input of the quantizer at the desired 200 MHz frequency and used for the calibration of the NTF. Over 25% variations on the loop parameters were intentionally introduced; this results in a notch's frequency around 250 MHz instead of 200 MHz. After several iterations using the aforementioned algorithm, the loop notch's frequency is tuned to the desired value by just monitoring the power of the test tone set at 200 MHz and adjusting the bank of capacitors used in the loop filter for that purpose. Figure 8.18b shows the ADC spectrum after calibration. The algorithm stops when the power of the tone at quantizer's output is at its minimum value; e.g. −65 dB at the output while the power of the test tone applied to the quantizer input is −10 dB. Once the loop's notch frequency is tuned, there is room for 3–8 dB of SQNR improvement by fine tuning DAC coefficients and excess loop delay.

8.5 Conclusion

OTA-C filters promise very good frequency response for high frequency applications; however they have limited linearity and are sensitive to process variations. To address the first issue, linearization methodologies for OTAs were discussed. Although linearity can be improved, most of the proposed techniques bring significant overhead in terms of power consumption, silicon area and noise. Furthermore, signal-to-distortion ratio improves at the cost of higher signal-to-noise ratio. Therefore, OTA linearization brings several trade-offs that needs to be carefully evaluated. The second issue, sensitivity of OTA-C filters to process variations, was

addressed using software-based calibration schemes. Application-specific solutions were demonstrated in Sections 8.3 and 8.4 showing that digital calibration methods can be employed to improve the accuracy of continuous-time filters.

References

1. R. Schreier, et al., "A 375-mW quadrature bandpass ΔΣ ADC with 8.5-MHz BW and 90-dB DR at 44 MHz," *IEEE J. Solid-State Circuits*, vol. 41, no. 12, pp. 2632–2640, Dec. 2006.
2. V. Dhanasekaran, M. Gambhir, M. M. Elsayed, E. Sanchez-Sinencio, J. Silva-Martinez, C. Mishra, L. Chen, E. Pankratz, "A 20 MHz BW 68 dB DR CT ΔΣ ADC based on a multi-bit time-domain quantizer and feedback element," *Proc. IEEE ISSCC*, pp. 174–175, Feb. 2009.
3. Y. Tsividis, "Continuous-time filters in telecommunication chips," *IEEE Communications Magazine*, vol. 163, no. 4, pp. 132–137, Apr. 2001.
4. J. Silva-Martínez, M. Steyaert and W. Sansen, *High Performance CMOS Continuous-time Filters*, Norwell, MA: Kluwer, 1993.
5. R. Schaumann and V. Valkenburg, *Design of Analog Filters*, New York: Oxford, 2001.
6. D. K. Shaeffer, et al., "A 115-mW, 0.5-μm CMOS GPS receiver with wide dynamic-range active filters," *IEEE J. Solid-State Circuits*, vol. 33, no. 12, pp. 2219–2231, Dec. 1998.
7. J. Silva-Martinez, J. Adut, J. M. Rocha-Perez, M. Robinson and S. Rokhsaz, "A 60-mW 200-MHz continuous-time seventh-order linear phase filter with on-chip automatic tuning system," *IEEE J. Solid-State Circuits*, vol. 38, no. 2, pp. 216–225, Feb. 2003.
8. Y. W. Choi and H. C. Luong, "A high-Q and wide-dynamic-range 70 MHz CMOS bandpass filter for wireless receivers," *IEEE Trans. Circuits Syst. II*, vol. 48, no. 5, pp. 433–440, May 2001.
9. S. Lindfors, K. Halonen and M. Ismail, "A 2.7-V elliptical MOSFET-only g_mC-OTA filter," *IEEE Trans. Circuits Syst. II*, vol. 47, no. 2, pp. 89–95, Feb. 2000.
10. A. Lewinski and J. Silva-Martinez, "OTA linearity enhancement technique for high frequency applications with IM3 below −65dB," *IEEE Trans. Circuits Syst. II*, vol. 51, no. 10, pp. 542–548, Oct. 2004.
11. F. Behbahani, et al., "A broad-band tunable CMOS channel-select filter for a low-IF wireless receiver," *IEEE J. Solid-State Circuits*, vol. 35, no. 4, pp. 476–489, Apr. 2000.
12. M. Chen, et al., "A 2-Vpp 80–200-MHz fourth-order continuous-time linear phase filter with automatic frequency tuning," *IEEE J. Solid-State Circuits*, vol. 38, pp. 1745–1749, Oct. 2003.
13. A. Wiesbauer, et al., "A fully integrated analog front-end macro for cable modem applications in 0.18-μm CMOS, *IEEE J. Solid-State Circuits*, vol. 37, pp. 866–873, Jul. 2002.
14. A. Yoshizawa and Y. P. Tsividis, "Anti-blocker design techniques for MOSFET-C filters for direct conversion receivers," *IEEE J. Solid-State Circuits*, vol. 37, pp. 357–364, Mar. 2002.
15. A. Lewinski et al., "A 30 MHz 5th-order elliptic low-pass CMOS filter with 65 dB spurious free dynamic range," *IEEE Trans. Circuits and Syst. I*, vol. 54, pp. 469–480, Mar. 2007.
16. J. Sevenhans and M. Van Paemel, "Novel CMOS linear OTA using feedback control on common source node," *Electronics Letters*, vol. 27, pp. 1873–1875, Sep. 1991.
17. P. R. Gray, P. J. Hurst, S. H. Lewis and R. G. Meyer, *Analysis and Design of Analog Integrated Circuits*, New York: Wiley, 2001.
18. T. W. Fischer, B. Kelleci, K. Shi, A. I. Karsilayan and E. Serpedin, "An analog approach to suppressing in-band narrowband interference in UWB receivers," *IEEE Trans. Circuits Syst. I*, vol. 54, no. 5, pp. 941–950, May 2007.
19. K. Shi, Y. Zhou, B. Kelleci, T. W. Fischer, E. Serpedin and A. I. Karsilayan, "Impacts of narrowband interference on OFDM-UWB receivers: analysis and mitigation," *IEEE Trans. Signal Processing*, vol. 55, no. 3, pp. 1118–1128, Mar. 2007.
20. P. Kallam, E. Sanchez-Sinencio and A. I. Karsilayan, "An enhanced adaptive Q-tuning scheme for a 100 MHz fully-symmetric OTA-based bandpass filter," *IEEE J. Solid-State Circuits*, vol. 38, no. 4, pp. 585–593, Apr. 2003.

21. G. Colli and F. Montecchi, “Low voltage low power CMOS four quadrant analog multiplier for neural network applications,” *Proc. IEEE ISCAS*, vol. 1, pp. 496–499, May 1996.
22. F. Silva-Rivas, C.Y. Lu, P. Kode, B. K. Thandri, and J. Silva-Martinez, “Digital based calibration technique for continuous-time bandpass sigma-delta analog-to-digital converters,” *Analog Integrated Circuits and Signal Processing*, vol. 59, pp. 91–95, Apr. 2009.

Chapter 9
Source-Follower-Based Continuous Time Analog Filters

Stefano D'Amico, Marcello De Matteis, and Andrea Baschirotto

Abstract Continuous-time (CT) analog filters based on source-follower circuits and on a local positive feedback to synthesize negative resistances and complex poles are here proposed. The intrinsic source-follower feedback allows these filters to perform large linearity for smaller $V_{OV}(= V_{GS} - V_{TH})$. This is exactly the opposite of the other CT filters where linearity performance improves with V_{OV} and with the current consumption increasing. Two circuit implementations will be shown. The first filter uses a cascade topologies to synthesize a fourth-order lowpass filter. In a 0.18 μm CMOS at 1.8-V supply, it achieves a 17.5 dBm IIP3 and a −40 dB HD3 for a 600-$mV_{pp-diff}$ input signal amplitude. A 24 μV_{rms} noise gives a DR = 79 dB, with 2.25 mA current consumption. The second filter exploits a ladder topology to synthesize a sixth order low-pass filter frequency response. In a 0.13 μm CMOS technology with $V_{DD} = 1.2$ V, the cut-off frequency is 280 MHz while the DC gain is 0 dB. An 11 dBm IIP3 has been measured. The output noise is about −140 dBm at 3 MHz.

9.1 Introduction

The growing request of wireless area networks requires a significant effort on low-power, low-cost, and highly integrated circuits, where the targets are CMOS technological scaling down, power consumption reduction and high data rate. The state-of-the-art of the telecom transceivers uses System-On-Chip implementations [1–6]. In SoC analog and digital circuits share the same die area, supply voltage and CMOS technology. Consequently, power consumption of portable devices is significantly reduced, saving battery life or releasing a significant amount

S. D'Amico (✉) and M. De Matteis
Department of Innovation Engineering, University of Salento, Lecce, Italy
e-mail: stefano.damico@unile.it,marcello.dematteis@unile.it

A. Baschirotto
Department of Physics "G. Occhialini", University of Milano Bicocca, Milano, Italy
e-mail: andrea.baschirotto@unimib.it

A.H.M. van Roermund et al. (eds.), *Analog Circuit Design: Smart Data Converters, Filters on Chip, Multimode Transmitters*, DOI 10.1007/978-90-481-3083-2_9,
© Springer Science+Business Media B.V. 2010

of power. This budget of additional power can be used to increase complexity of digital circuits and consequently to improve the overall receiver performance. The most recent transceivers characteristics demonstrate the analog circuits like low pass filters (LPF) and programmable gain amplifiers (PGA) embedded in the base band chain of the telecom receivers, are responsible of a considerable amount of power with respect to the entire system [7–14]. The case of the baseband continuous-time Low-Pass Filter (LPF) in base band chain fully-integrated direct-conversion receivers is here addressed. Selection filters are needed in order to separate the base band input signal from the adjacent unwanted channels. Anti-aliasing for the following ADC and out-of-band frequencies rejection are then other important tasks for analog filters. In addition the filter have to perform high linearity and wide bandwidth to meet the requirements of telecom systems in terms of SNR and bit rate.

Among several possible solutions, Active-RC [7, 8] and Active-g_m-*RC* [11, 12] filters guarantee high linearity but with a larger power consumption (due to the closed loop topology) than with g_m-*C* [9] filters that, however, suffer from reduced linearity. Moreover in Active-RC and Active-g_m-*RC* filters, resistors are needed, and consequently the noise power increases, while g_m-*C* filters save the noise power but require large overdrive voltage, in order to comply with severe telecom linearity requirements. That means to increase power consumption and to reduce the filter maximum allowable output swing.

In source follower circuits no resistors are needed, and large g_m is obtained reducing the overdrive voltage and the power consumption, while maintaining linearity due to its intrinsic feedback. Moreover, large g_m means low MOS transistors noise power. From this point of view, source follower appears a good trade-off between Active RC topologies and g_m-*C* filters. In this paper the source follower circuit has been investigated in order to synthesize complex poles, needed for high order analog filters.

The paper is organized as follows. Section 9.2 introduces some important aspects of the source follower circuits, in terms of transfer function, linearity and signal swing. Then in Section 9.3 a fourth order CT analog filter is presented. The filter is composed by the cascade of two source-follower-based cells. Each cell synthesizes two complex poles. Positive feedback is exploited in order to synthesize one complex poles pair. This filter satisfies the WLAN 802.11.a/b/g baseband specifications. The first part of Section 9.3 illustrates the basic characteristics of the filter in terms of transfer function, noise and linearity performance. The minimum supply voltage requirements for the filter are then reported and the quality factor sensitivity function of the biquadratic cell is calculated. A prototype has been integrated in 0.13 μm CMOS technology with 1.2 V supply voltage. The second design is presented in Section 9.4. It synthesizes with a ladder structure a sixth-order low-pass frequency response and has been integrated in a 0.13 μm CMOS technology with $V_{DD} = 1.2$ V. By alternation of negative and positive resistances it is possible to implement a ladder net where the negative resistances allows to synthesize complex poles. The filter complies with the requirements of the base band chain for a low data-rate (LDR) UWB receiver.

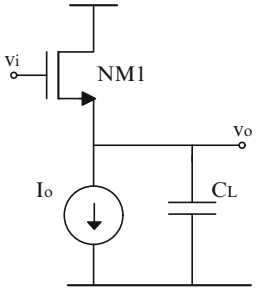


Fig. 9.1 Source follower circuit

9.2 Source-Follower Circuit

The filter here proposed is based on a complex source-follower topology [14].

The source-follower, shown in Fig. 9.1, is the basic cell of the proposed second-order cell. The source-follower transfer function can be written as:

$$G(s) = \frac{g_{m1}}{g_{m1} + g_{ds1} + g_{d0}} \cdot \frac{1}{1 + s \cdot \frac{C_L}{g_{m1}+g_{ds1}+g_{d0}}} \tag{9.1}$$

where g_{ds1} and g_{m1} are the NM1 output conductance and transconductance respectively, g_{d0} is the current source I_o output conductance. The source-follower presents some interesting features that make it very attractive to design analog filters. First, it has a good in-band linearity, due to the presence of the local feedback. As any feedback structure, its linearity improves with a large closed-loop gain. The loop gain is given by:

$$G_{loop}(s) = \frac{g_{m1}}{g_{ds1} + g_{d0}} \cdot \frac{1}{1 + s \cdot \frac{C}{g_{ds1}+g_{d0}}} \tag{9.2}$$

while the v_{gs}/v_{in} (gate-source voltage of the NM1 transistor) can be written as:

$$\frac{v_{gs}}{v_{in}}(s) = \frac{v_{in} - v_{out}}{v_{in}}(s) = \frac{g_{m1} + g_{ds1}}{g_{m1} + g_{ds1} + g_{d0}} \cdot \frac{1 + s \cdot \frac{C}{g_{ds1}+g_{d0}}}{1 + s \cdot \frac{C}{g_{m1}+g_{ds1}+g_{d0}}} \tag{9.3}$$

Figure 9.2 shows the frequency response of Eqs. 9.1, 9.2 and 9.3. G(s), $G_{loop}(s)$ and $V_{gs}/V_{in}(s)$ has been obtained without considering the parasitic poles and zeros, present in source-follower circuit and whose frequency is typically very much larger than the baseband applications.

The values of the NM1 and I_o transconductance and output resistance are reported in Table 9.1. In this design example, the source follower circuit is a first order low-pass filter, where its −3 dB frequency is placed at 9.5 MHz. The dc gain is −1.5 dB, and the drop at low frequency with respect to the ideal 0 dB gain depends on the finite values of NM1 g_m and g_{ds}, and g_{d0}.

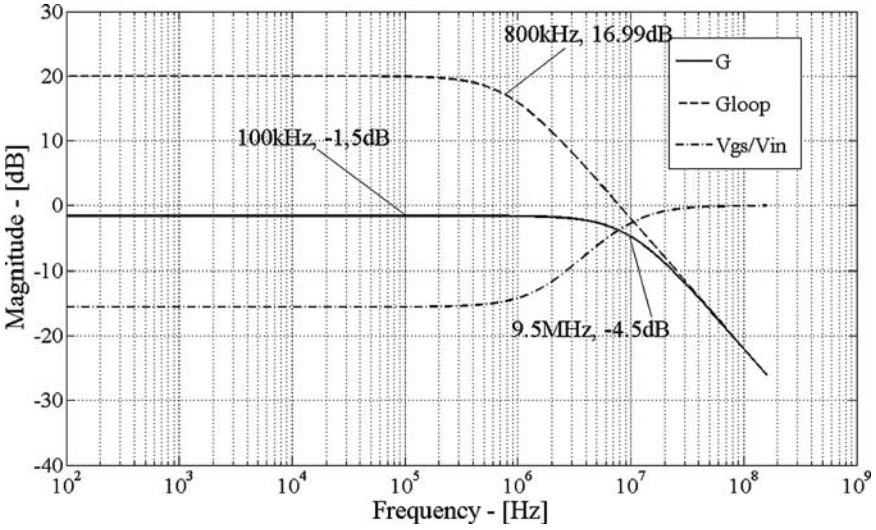


Fig. 9.2 Source follower gain, G_{loop} and V_{gs}/V_{in} vs. frequency

Table 9.1 Source follower small signal parameters

Small signal parameter	Value
NM1 – g_m	1 mA/V
NM1 – g_{ds1}	0.1 mA/V
I_o Output Conductance – g_{d0}	0.1 mA/V

At low frequencies a larger transconductance ($g_m = 2{\cdot}I/V_{ov}$) value is achieved for lower V_{ov}. This results in a larger loop gain and then in a better linearity. The v_{gs}/v_{in} frequency response confirms these considerations, showing the reduced drop of the in-band gate-source signal swing. This basic conclusion completely differs from other active filters (like g_m-C filters, for instance), where the linearity is improved at the cost of larger V_{ov}, and then larger current (for a given g_m) and power consumption. Breaking the dependence V_{ov}-versus-linearity immediately has a large impact on the power performance. Minimizing V_{ov} corresponds to reduce the current level to achieve the same g_m value. This is reflected in a substantial power saving for the same linearity level. The above concept corresponds to the fact that the source-follower processes the signal directly in the voltage domain, avoiding having to convert it into a current and then back to voltage. In this way, it is possible to increase the linearity range, as the main source of distortion in g_m-C filters comes from the conversion of the voltage signal into a current. This operation is performed by the transconductors in a g_m-C filter. Each filter transconductor processes large signals in a open-loop configuration; therefore, they introduce distortion. Moreover, a time constant is obtained depending only on the transistors transconductance and on the capacitive load. This means that the circuit does not need to drive any resistive load, avoiding to consume current under signal regime, I_{swing}. This current can

be evaluated as follows:

$$I_{swing} = \frac{V_{swing}}{R_L} \tag{9.4}$$

where V_{swing} is the voltage signal swing and R_L is the eventual resistive load from the output node to the ground. This current, which corresponds to the dynamic power consumption, is provided by the filter active device, therefore increasing its static current requirement. The absence of a resistive load to be driven gives higher power efficiency to the source-follower.

It is worth to clarify that these considerations are strongly related with the $G_{loop}(s)$ behavior reported in Fig. 9.2. While, at low frequency, gate-source voltage swing is minimized by the loop gain, close to the pole frequency and higher, the loop gain decreases, maximizing the gate-source voltage swing. Out-of-band linearity is expected to be critical for this kind of circuits.

Moreover, the source follower circuit presents more advantages that allow reduction of the current consumption requirement. For example, no circuital parasitic poles are present. In fact, each circuit node corresponds to a pole included in the cell transfer function. Therefore, there is no need of extra current to push parasitic poles to higher frequencies. The output common-mode voltage is self-biased by the transistor NM1, without adding any additional circuit, as the common-mode feedback circuit. In addition, the source-follower can drive a resistive load without substantially modifying its linearity performance and its pole frequency.

9.3 A Source-Follower-Based Cascade CT Filter

The advantages of the source-follower are reported in the second-order low-pass cell shown schematically in Fig. 9.3 [14]. This cell presents an optimized single-branch fully differential structure and operates like a "composite" source-follower.

All the transistors are designed with the same sizes and draw the same current levels. As a consequence, they all exhibit the same transconductance, which can be written as

$$g_{m1} = g_{m2} = g_{m3} = g_{m4} = g_m \tag{9.5}$$

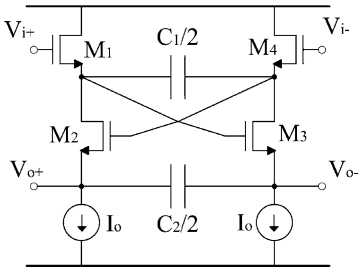


Fig. 9.3 Proposed biquadratic cell

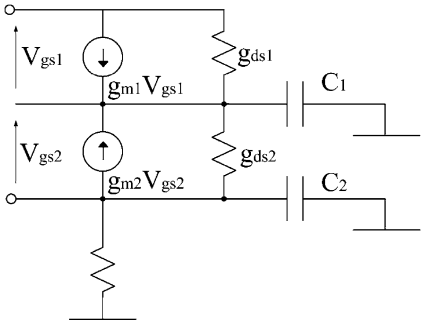


Fig. 9.4 Half small-signal equivalent circuit for the differential mode

Thus, it performs an ideal unitary DC-gain. The key proposal of this cell is the positive feedback in MOS devices M_2 and M_3, which allow the synthesis of two complex poles. Figure 9.4 shows the half small-signal equivalent circuit, valid for the differential mode. The stability of the positive feedback is guaranteed, in fact, the loop gain obtained by cutting the loop differentially at the gates of and at low frequency is given by

$$G_{loop}(s) = \frac{g_{m2} \cdot (g_{d0} + g_{ds2})}{(g_{m1} + g_{ds1}) \cdot (g_{m2} + g_{d0} + g_{ds2})} \tag{9.6}$$

therefore, it is always lower than 1.

Assuming that the transistor's output conductance is much smaller than the transconductance, the filter transfer function is:

$$H(s) = \frac{1}{s^2 \cdot \frac{C_1 \cdot C_2}{g_m^2} + s \cdot \frac{C_1}{g_m} + 1} \tag{9.7}$$

The filter parameters (the pole frequency ω_o, the quality factor Q, and the DC-gain K) are given by:

$$\begin{cases} \omega_o = 2\pi f_0 = \frac{g_m}{\sqrt{C_1 \cdot C_2}} \\ Q = \sqrt{\frac{C_2}{C_2}} \\ |K| = 1 \end{cases} \tag{9.8}$$

In the case that the MOS output conductance is not negligible with respect to the transconductance, the transfer function and the relative parameters are modified as:

$$H(s) = -\frac{\frac{g_{m2} \cdot g_{m1}}{g_{ds0} \cdot g_{m1} + g_{ds0} \cdot g_{ds1} - g_{ds1} \cdot g_{m2} + g_{m1} \cdot g_{m2}}}{s^2 \cdot \frac{C_1 C_2}{\substack{g_{ds0} \cdot g_{m1} + g_{ds0} \cdot g_{ds1} \\ -g_{ds1} \cdot g_{m2} + g_{m1} \cdot g_{m2}}} + s \cdot \frac{\substack{C_2 \cdot g_{m1} + C_2 \cdot g_{ds0} - C_2 \cdot g_{m2} + C_1 \cdot g_{ds1} + C_1 \cdot g_{m2} \\ g_{ds0} \cdot g_{m1} + g_{ds0} \cdot g_{ds1} \\ -g_{ds1} \cdot g_{m2} + g_{m1} \cdot g_{m2}}}{\substack{g_{ds0} \cdot g_{m1} + g_{ds0} \cdot g_{ds1} \\ -g_{ds1} \cdot g_{m2} + g_{m1} \cdot g_{m2}}} + 1} \tag{9.9}$$

However, since the technology scaling results in an excess of transition frequency (f_T) for the baseband applications (such as filters, etc.), a design guideline is to

increase the channel length size (L) in order to reduce transfer function deviations. This L size increase only reduces the transistor f_T, which in any case remains sufficiently large for the baseband applications. The choice of larger L also improves transistor matching performance [15] and the robustness of the filter with respect to the output MOS conductances variations. In fact, the sensitivity of the Q factor with respect to the output MOS conductance, g_{ds1}, can be estimated as follows:

$$S_{Q,gds1} = \frac{\partial Q}{\partial g_{ds1}} \cdot \frac{g_{ds1}}{Q} \cong -\frac{3}{2}\frac{g_{ds1}}{g_{m1}} \tag{9.10}$$

As for the basic source-follower cell, the above second-order cell features these advantages which can be exploited to reduce the power consumption:

- No circuital parasitic poles are present, disabling the power cost of pushing non-dominant singularities at high frequency.
- No CMFB circuit: the output CM voltage is fixed to be lower than the input CM voltage.
- Low output impedance: as the source-follower, the filter can drive a resistive load with negligible effects on the filter performance in terms of linearity and filter transfer function accuracy.

9.3.1 Linearity Performance

As the source-follower, this structure presents the key advantages of a large linear range. Assuming that the third-harmonic-distortion (HD3) is the main contribution to the total harmonic distortion because the structure is fully differential, at low frequency and assuming that transistors are pushed to work in weak inversion, the HD3 can be evaluated as follows:

$$HD3 = \frac{1}{6} \cdot \frac{v_{in}^2}{\left(n \cdot \frac{kT}{q}\right)} \cdot \frac{1}{\left(1 + \frac{g_m}{g_{d0}}\right)^2} \cong \frac{1}{6} \cdot \frac{v_{in}^2}{\frac{1}{\lambda^2}} \tag{9.11}$$

where n is the slope factor, and λ is the channel modulation coefficient [16]. From Eq. 9.11, the linearity improves by increasing λ, which implies the use of longer transistors, as it is the case of transfer function accuracy (see Eq. 9.10). This formula is valid until transistors keeps on working in weak inversion. In fact the linearity is fixed by the weak inversion-to-linear borderline of M2, which allows a swing about equal to V_{TH}, i.e., it depends on the technology choice. Since the target is operating with high g_m, despite the low V_{ov}, the structure can be implemented with large linearity also with BJT devices, as shown in Fig. 9.5.

In this case, the linearity range is about V_{be}. At the same time, due to the high of bipolar transistors, this solution would be particularly power efficient.

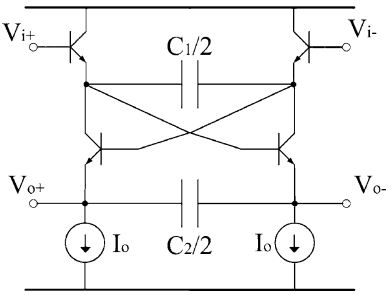


Fig. 9.5 BJT cell version

9.3.2 Noise Performance

Due to the large bandwidth of the WLAN signal to be processed by this cell, the transistors thermal noise is dominant, and the input-referred noise spectral density at low frequency is

$$IRN^2 = \frac{64}{3} \cdot \frac{kT}{g_m} \tag{9.12}$$

This gives an integrated input-referred noise given by

$$V_{in,noise}^2 = \frac{64}{3} \cdot \frac{kT}{\sqrt{C_1 C_2}} \tag{9.13}$$

As expected, the above depends on the average value of the capacitances. From Eqs. 9.11 and 9.13, the dynamic range (DR) is calculated and can be expressed as follows:

$$IRN^2 = \frac{32}{3} \cdot \frac{V_{th}^2}{kT} \cdot \sqrt{C_1 \cdot C_2} \tag{9.14}$$

9.3.3 DC-Gain Loss

The proposed cell exhibits a DC-gain sensitivity to the bulk transconductance, whose effect is a DC-gain reduction. This effect in the simulation environment is well modeled, and the DC-gain can be calculated as a function of the ratio between the transistors bulk transconductance and their transconductance $\eta = g_{mb}/g_m$ as follows:

$$DC - gain = \frac{1-\eta}{1-\eta+\eta^2} \tag{9.15}$$

This effect can be reduced with a source–bulk connection. For the PMOS devices, the bulk connection is always available and this allows to avoid this gain loss. On the other hand, for the NMOS devices, this is possible if the technology has the double-well option. In this case, however, the parasitic capacitance between well

and substrate has to be taken into account. In the filter design proposed in the next section, the ratio η between the NMOS transistors bulk transconductance and their transconductance is 0.49. According to Eq. 9.15, the gain loss is 3.5 dB.

9.3.4 Minimum Supply Voltage

Supposing that the previous stage presents at least a $V_{DD} - V_{sat}$ (where $V_{sat} = V_{gs} - V_{th}$) output common mode, the minimum required supply voltage is:

$$\mathrm{V_{DD,min}} = \mathrm{V_{sat}} + \mathrm{V_{GS1}} + \mathrm{V_{GS3}} + \mathrm{V_{swing}} + \mathrm{V_{sat}} = 4\mathrm{V_{sat}} + 2\mathrm{V_{th}} + \mathrm{V_{swing}} \quad (9.16)$$

For a 0.18 µm CMOS technology, assuming $V_{sat} = 150\,mV$, $V_{GS1} = V_{GS3} = 550\,mV$, $V_{swing} = V_{th} = 400\,m$, a $V_{DD,min} = 1.8\,V$ is needed. A popular solution to reduce the $V_{dd,min}$ of stacked structures is to pass to a folded structure. The folded structure of the proposed cell is shown in Fig. 9.6. This structure requires a $V_{DD,min}$ slightly lower than the stacked one, and it is given by:

$$\mathrm{V_{DD,min}} = \mathrm{V_{sat}} + \mathrm{V_{GS1}} + \mathrm{V_{swing}} + \mathrm{V_{sat}} = 3\mathrm{V_{sat}} + 2\mathrm{V_{th}} + \mathrm{V_{swing}} \quad (9.17)$$

Using the above bias point for the devices, the $V_{DD,min}$ for the folded solution is 1.25 V. It, however, consumes double power because two branches are needed, and this motivates the preference for the single branch one, which is the more power optimized. Nonetheless, for low-voltage application, the folded topology would have to be considered.

9.3.5 Silicon Prototype Experimental Results: A Fourth-Order Cascade SFB CT Filter for WLAN Receivers

The proposed source-follower-based biquad cell has been validated by designing a benchmark prototype. A fourth-order Bessel low-pass filter satisfying typical

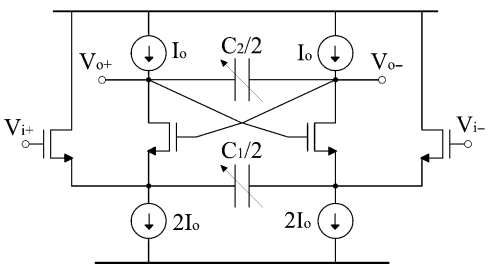


Fig. 9.6 Folded cell version

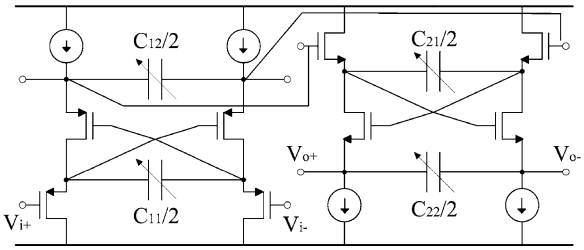


Fig. 9.7 Fourth-order filter schematic

Table 9.2 Basic parameters of each biquadratic cell

	C_1(pF)	C_2(pF)	g_m(mA/V)	Q	f_o(MHz)
Cell1	85.3	25.5	1.8	0.52	15.04
Cell2	71.4	50.1	1.8	0.81	16.86

WLAN 802.11.a/b/g receiver specifications has been designed as a cascade of two single-branch cells. Figure 9.7 shows the filter schematic and Table 9.2 the basic parameters of each biquadratic cell.

The first cell is made up of PMOS transistors and the second one is made up of nMOS transistors. The use of cascading PMOS and NMOS structures allows for compensation of the input-to-output common-mode voltage difference, typical of the source-follower and of the proposed source-follower-based second-order cell. All MOS devices have been designed with a 0.5 μm channel length to reduce output impedance effects on the frequency response accuracy and to improve the linearity (as previously described). In any case, wider MOS sizes will increase the parasitic capacitances at each node. This can affect the capacitance matching and, thus, the Q. Therefore, this channel length design is a tradeoff between linearity and frequency response accuracy. All capacitors are realized with digitally controlled arrays. In this way, the filter cut-off frequency is then programmable in the 40% range with a 4-bit word, compensating technology, temperature and parasitic capacitance effects. This means that the tuning is independent on bias current, therefore V_{ov}, loop gain, and linear range are constant.

The prototype has been realized in a 0.18-μm CMOS technology. Figure 9.8 shows the chip microphotograph. In the device, two filters (for I&Q signal processing) have been realized. For the proposed solution, only 0.52 mm^2 are needed for a two-filter structure (I&Q topology). This means 0.26 mm^2 for each filter. The area occupation has been limited by using high-density capacitors (4 fF/mm^2). The measured transfer function is shown in Fig. 9.9. The filter transfer function exhibits a 10 MHz cut-off frequency and a −3.5 dB DC-gain. While the gain of the first cell is close to one, the gain of the second cell (NMOS-based) is affected by the bulk transconductance, as expected. In fact, in the adopted process no double well is available, disabling the cancellation of the NMOS bulk transconductance.

Figure 9.10 shows the in-band IM3 for two tones at 3 and 4 MHz of 100 mV each at the outputs. This corresponds to an in-band IIP3 of 17.5 dBm, as shown in

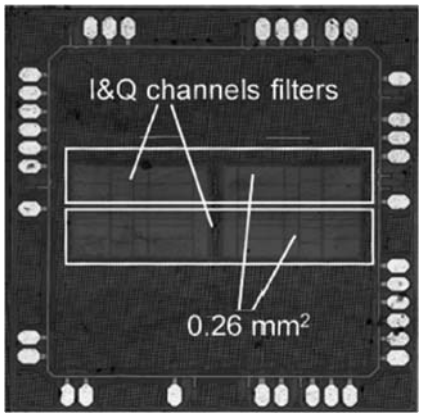


Fig. 9.8 Chip microphotograph

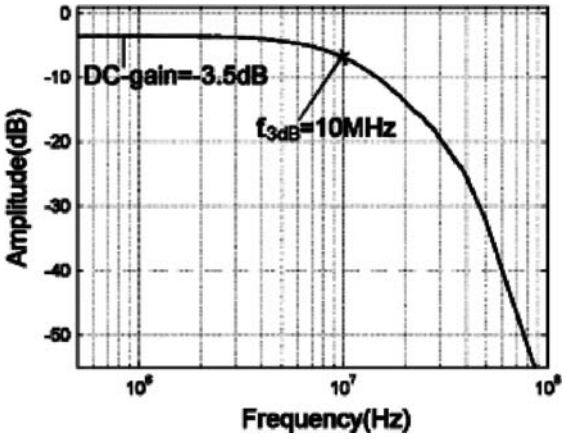


Fig. 9.9 Filter transfer function

Fig. 9.11. This test was repeated by using two tones spaced by 1 MHz and changing their central frequency. In Fig. 9.10, a graph plotting the IIP3 as a function of the two tones central frequency is reported. The minimum IIP3 is 12 dBm and it is obtained around the cut-off frequency. This depends on the local feedback gain reduction due to the presence of a capacitive load on the transistors sources.

The 1 dBcp, measured by using a 1 MHz input tone, is equal to 5 dBm (i.e., 1.15 V from a 1.8-V supply). In Fig. 9.12, the output power is plotted as a function of the input power when a 1 MHz sine is applied at the input. From these measurements, the 1 dBcp is obtained. In Fig. 9.13, the linearity has been evaluated also in terms of HD3, plotted as a function of the input amplitude at 3 MHz. The HD3 is 40 dB for a 600 $mV_{pp.diff}$ input signal amplitude. As evident from Fig. 9.13, there is a good agreement between the measured results and the analytical model prediction given by Eq. 9.11. A summary of the filter perfomance is given in Table 9.3.

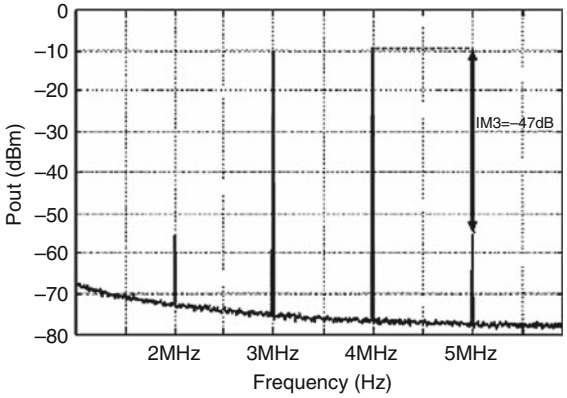
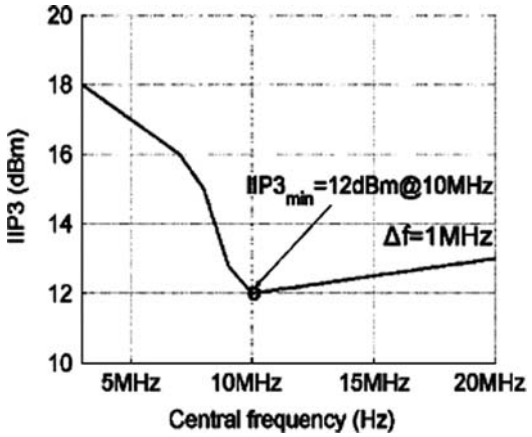


Fig. 9.10 Output power vs. frequency for two tones test

Fig. 9.11 IIP3 vs. the central frequency of the two tones



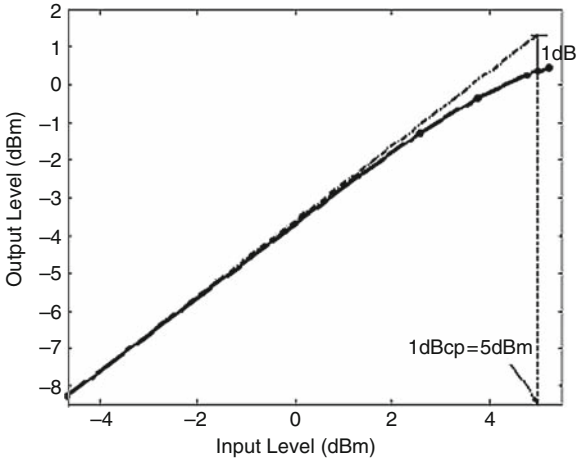


Fig. 9.12 Output power vs. input power for a 1 MHz input sine

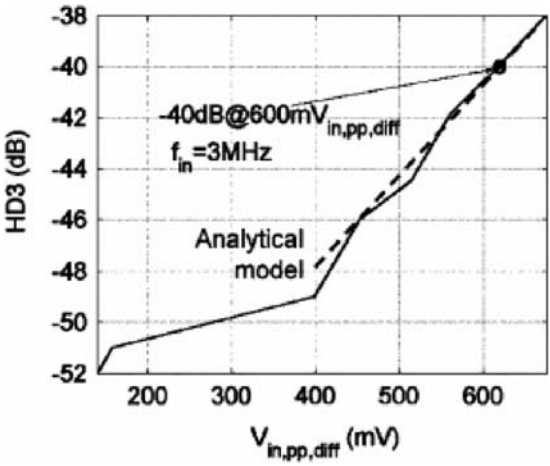


Fig. 9.13 HD3 vs. input sine amplitude @ 3 MHz

Table 9.3 Filter performance summary

Technology	CMOS 0.18 μm
Die area	0.26 $mm^2 \times 2$
Power supply	1.8 V
Current consumption	2.28 mA
Power consumption	4.1 mW
DC-gain	−3.5 dB
f_{-3} dB	10 MHz
f_{-3} dB Tuning range	±40%
IRN	7.5 nV/$\sqrt{}$Hz
$V_{in,noise}$	24 μV_{rms}
DR (HD3 = −40 dB)	79 dB
in-band IIP3 ($f_1 = 3$ MHz, $f_2 = 4$ MHz)	17.5 dBm
1 dB cp	5 dBm
HD3 (600 mV_{pp} @ 3 MHz)	−40 dB

9.4 A Source-Follower-Based Ladder CT Filter

In this section the source-follower technique is extended to the design of single-loop high-order source-follower-based continuous-time filters that, as for the ladder filters, would present lower amplitude frequency response sensitivity to component value variations. An efficient CMOS realization validates the proposal, which requires lower power and lower area with respect to other solutions. The proposed filter architecture is based on two, positive and negative, first-order building blocks as shown in Fig. 9.14. In the negative block, R^* is not a standard impedance and its small-signal voltage-to-current relationship is:

$$i = -(v_i + v_o)/R \tag{9.18}$$

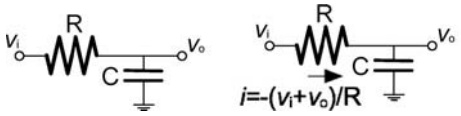


Fig. 9.14 First-order positive (*left*) and negative (*right*) building blocks

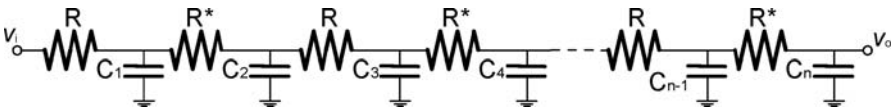


Fig. 9.15 Generalized filter architecture

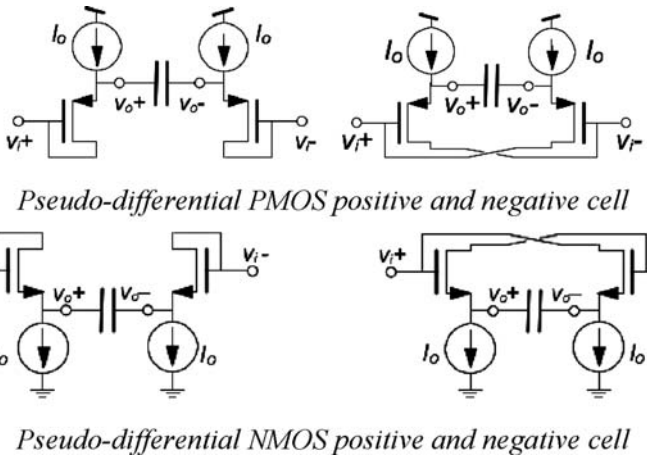


Fig. 9.16 Source-follower first-order building blocks. Pseudo-differential PMOS positive and negative cell, pseudo-differential NMOS positive and negative cell

Composing a sequence of these cells allows synthesizing high-order filters. The sequence can be the regular alternation of positive and negative cells or they can be connected with a different order. A negative cell is needed to synthesize complex poles. In the following the regular alternation is considered and results in the generalized filter architecture of Fig. 9.15, whose general transfer function is:

$$H(s) = \frac{1}{\left[s^n R^n \cdot \prod_{1,\ldots,n} C_i\right] + \left[s^{n-1} R^{n-1} \cdot \prod_{1,\ldots,n-1} C_i\right] + \ldots + 1} \tag{9.19}$$

This is the transfer function of a low-pass filter with unitary DC-gain. The advantage of this approach is significant in conjunction with the proposed efficient CMOS implementation shown in Fig. 9.16, for the positive and negative building blocks with PMOS devices (NMOS can also be considered). They are in pseudo-differential structure, as they will be in the final design. Designing a filter in this

way guarantees to the structure the above advantages of the source follower. In addition, the generic time-constant is defined as:

$$\tau = \frac{C}{g_m} = \frac{C \cdot V_{ov}}{2 \cdot I} \tag{9.20}$$

9.4.1 Filter Circuital Topology

A particular discussion regards the input stage. The overall filter can be built-up with the cells above presented. In this way, the input impedance depends on the filter order and on the input signal frequency. At low frequency the input impedance is equal to R for an even-order structure, while it is equal to infinite (ideally) for an even-order structure.

On the other hand, at high frequency for any filter order the input impedance is equal to R. If this impedance level is critical for the filter under development, just the first cell can be replaced by an ideal source-follower. In this case the input impedance is given by the source-follower input impedance that is quite large.

9.4.2 Minimum Supply Voltage

These cells can operate with a theoretical minimum supply (V_{DDmin}):

$$\mathrm{V_{DDmin}} = 3\mathrm{V_{ov}} + \mathrm{V_{th}} + \mathrm{V_{swing}} \tag{9.21}$$

where V_{th} is the threshold voltage, while V_{swing} is the signal swing. From Eq. 9.21, for a 0.13 µm CMOS technology, −31 dB THD at $V_{swing} = 255\,\mathrm{mV}$, $V_{th} = 300\,\mathrm{mV}$ and $V_{ov} = 100\,\mathrm{mV}$ overdrive, $V_{DDmin} = 855\,\mathrm{mV}$ results.

9.4.3 Silicon Prototype Experimental Results: A Sixth-Order Ladder SFB CT Filter for UWB Receivers

A sixth-order low-pass filter prototype in a 0.13 µm CMOS technology with $V_{DD} = 1.2\,\mathrm{V}$ validates the proposal. The filter schematic is shown in Fig. 9.17. The filter performance (Table 9.5) satisfies the requirements of the channel selection filter for a LDR-UWB receiver. The critical issue of source-follower-based cells (i.e. the dc voltage drop equal to V_{GS} between input and output nodes) is here solved by alternating NMOS and PMOS cells. In this way, the DC level is restored. Figure 9.18 shows the filter transfer function. The cut-off frequency is 280 MHz while the DC-gain is about 0 dB. The linearity has been evaluated in term of in-band and

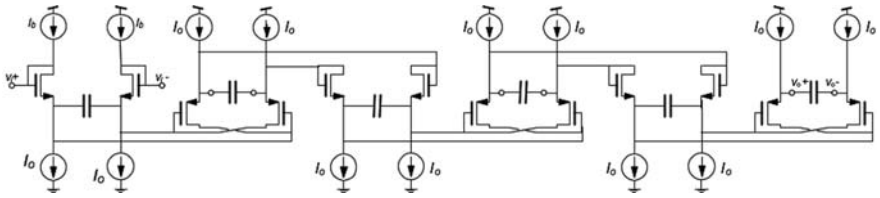


Fig. 9.17 Overall filter schematic

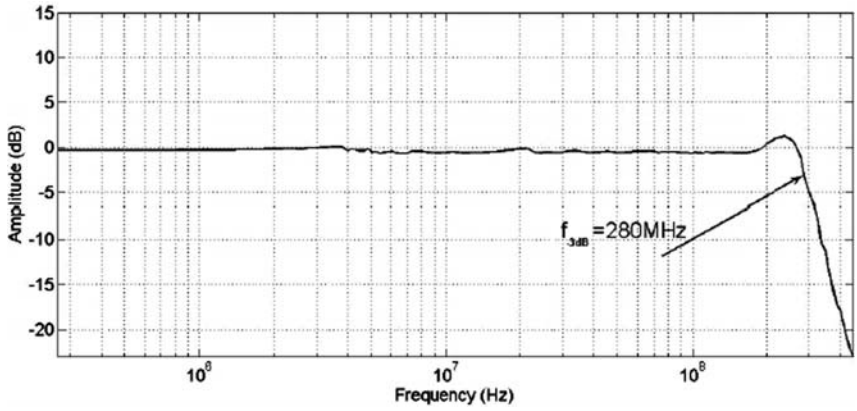


Fig. 9.18 Filter frequency response

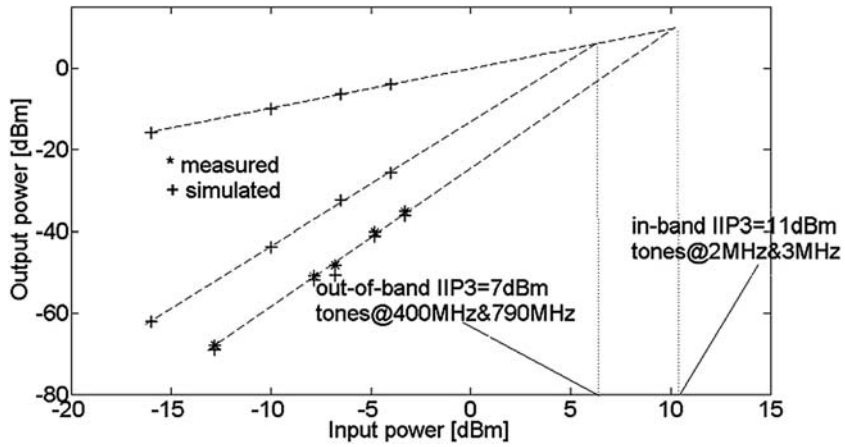


Fig. 9.19 In-band/out-of-band IIP3

out-of-band IIP3 as shown in Fig. 9.19. A 11 dBm in-band IIP3 have been measured, while the out-of-band IIP3 is about 7 dBm. The in-band/out-of-band linearity tests are obtained by testing the filter in two different conditions, as reported in Table 9.4.

Table 9.4 Linearity test setting

Linearity test	Tones frequency
In-band	2 and 3 MHz
Out-band	400 and 790 MHz

Table 9.5 UWB source-follower-based filter. Performance summary

	This design
V_{DD}	1.2 V
CMOS technology	0.13 μm
Current consumption	100 μA
Power consumption	0.12 mW
Filter order	Sixth
G	0 dB
f_{-3} dB	280 MHz
In-band IIP3	11 dBm
Out-of-band IIP3	7 dBm
1 dB cP	−1.2 dBm
IRN (10 ÷ 250 MHz)	22 nV/$\sqrt{}$Hz
Circuit area	0.0 18 mm^2
THD (v_{in} = 400 mV_{pp}@1 MHz)	40 dBc

Out-of-band linearity is one of the most important requirements for the UWB filters, due to the presence of strong out-of-band interferers at the input of the receivers. The basic building block of the analog filter here presented is a source follower circuit. Large gate source voltage swing is present at high frequency (see Fig. 9.2), in particular in the filter stop band and this results in worst out-of-band linearity performance with respect to the in-band signal behavior. In fact considering the out-of-band linearity test, where two interferers are placed at 400 MHz and 790 MHz, the lowest frequency inter-modulation product is at 10 MHz, which is an in-band signal for the filter. Furthermore the 400 MHz tone is placed close to the filter cut-off frequency. That can be problematic because, looking at Fig. 9.2, where a very simple source follower circuits has been considered, the maximum value of the gate-source voltage swing occurs very close to the pole frequency.

Anyway, the source-follower-based circuits object of this work, suffer from out-of-band linearity, due to the large signal swing, but improve the linearity performance with respect to the single pole standard source follower circuits. In fact, the presence of complex poles performs a considerable amount of filtering of the out-of-band interferers, and this is more and more important in telecom systems. For this reason no significant degradation of the out-of-band IIP3 is observed in the fourth order WLAN filter (see Fig. 9.11) and in the sixth order UWB filter (see Fig. 9.19), with respect to the in-band IIP3.

A possible approach in order to improve the out-of-band linearity performance in this filter is to increase the overdrive voltage, and consequently to set properly the C_i capacitance in order to maintain the same frequency response (see Eq. 9.20).

In this prototype, the total current is about 100 μA for a total power supply of 120 μW, which is favorable compared with the state-of-the-art of the UWB receivers. The output noise is −140 dBm @ 3 MHz. It is basically thermal noise, which is widely dominant in 250 MHz bandwidth range. Figure 9.20 shows the chip microphotograph. The chip area occupation is as small as $200 \times 90\ \mu m^2$. In this filter the total amount of capacitance is 500 fF, which occupies 0.01 mm^2.

The filter can be favorable compared with other continuous time analog filters with similar −3 dB frequency values, as shown in Table 9.6. These filters can be compared in terms of the Figure-of-Merit calculated in Eq. 9.22 and shown in Figure 9.21.

$$FOM = \frac{PW}{8kT \cdot N \cdot DR \cdot f_{-3dB}} \tag{9.22}$$

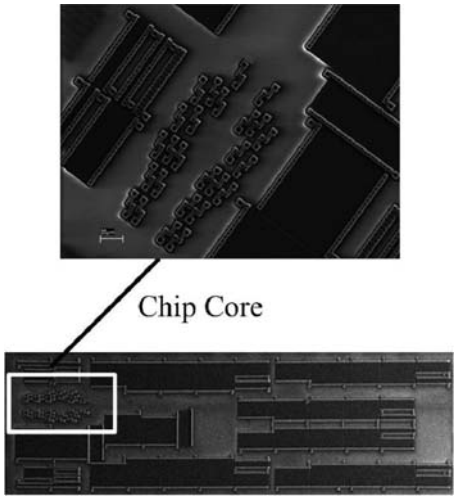


Fig. 9.20 Chip microphotograph

Table 9.6 Filter comparison

	[17]	[18]	[19]	[20]	[21]	This Work
V_{DD}[V]	3	2.3	2	2.5	1.2	1.2
CMOS	0.35 μm	0.35 μm	0.25 μm	0.25 μm	0.13 μm	0.13 μm
Power cons. (mW)	60	72	216	120	24	0.12
Filter order	Seventh	Fourth	Seventh	Eighth	Sixth	Sixth
G (dB)	0	35	0	6 ÷ 14	12 ÷ 47	0
f_{-3}dB	200	200	150	120	240	280
DR (dB) (THD > 40 dBc)	50	52	65	45	67	50
Area	0.18 mm^2	0.025 mm^2	–	0.25 mm^2		0.018 mm^2

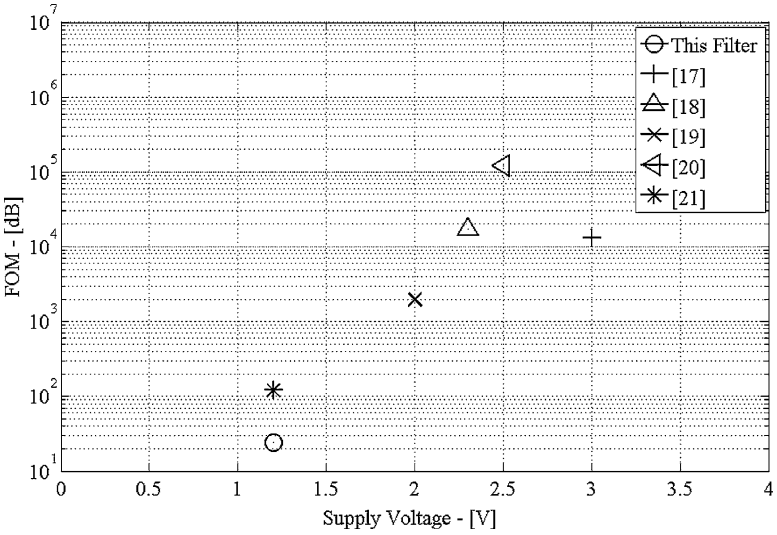


Fig. 9.21 Figure-of-merit comparison

9.5 Conclusions

A design technique for analog filters based on source follower circuit has been presented. Two different circuital topologies have been reported in this paper. The design approach has been validated by the experimental results of two filters for WLAN and UWB base band chain systems, respectively.

The filters based on source follower present the following key advantages. A large linearity is achieved for a low overdrive. This, in conjunction with the possibility of realizing a full biquadratic cell in a single branch, corresponds to a low power consumption for a given pole frequency (as demonstrated by the comparison with other filters). In addition, low output impedance is achieved and no common-mode feedback is needed. Using a positive feedback, complex poles can be synthesized.

The first proposal is based on a single branch biquadratic cell. This cell is exploited in order to synthesize one complex poles pair. As a design example, in a 0.18 μm CMOS at 1.8 V supply, a fourth order 10 MHz filter for WLAN applications has been designed and fully characterized. It performs at 17.5 dBm IIP3 and −40 dB HD3 for a 600 mV input signal amplitude. A 24 μV_{rms} noise gives a 79 dB DR, with 2.25-mA current consumption. In the second design the source follower circuit has been used to synthesize a ladder network, composed by the regular alternation of positive and negative cells. The filter has been integrated in CMOS 0.13 μm technology and dissipates only 0.12 mW from a 1.2 V single supply voltage. The linearity measurements give 11 and 7 dBm of in-band and out-of-band IIP3 respectively.

Acknowledgments This research has been partially supported by the Italian National Program PRIN 2005, "Enabling blocks for the integration in CMOS technology of a Multi-Band OFDM "Ultra Wide Band" transceiver".

References

1. G. Hueber, et al., "A Single-Chip Dual-Band CDMA2000 Transceiver in 0.13 μm CMOS". Solid-State Circuits Conference, 2007. ISSCC Digest of Technical Papers. IEEE International. 11–15 Feb. 2007, pp. 342–607.
2. M. Simon, et al., "An 802.11a/b/g RF Transceiver in an SoC". Solid-State Circuits Conference, 2007. ISSCC. Digest of Technical Papers. IEEE International. 11–15 Feb. 2007, pp. 562–622.
3. T. Bernard, et al., "Single-Chip Tri-Band WCDMA/HSDPA Transceiver without External SAW Filters and with Integrated TX Power Control" Solid-State Circuits Conference, 2008. ISSCC Digest of Technical Papers. IEEE International. Feb. 2008, pp. 202–203.
4. R.B. Staszewski, et al., "A 24 mm^2 Quad-Band Single-Chip GSM Radio with Transmitter Calibration in 90 nm Digital CMOS". Solid-State Circuits Conference, 2008. ISSCC Digest of Technical Papers, IEEE International. Feb. 2008, pp. 208–209.
5. C. Namjun. et al., "A 60 kb/s-to-10 Mb/s 0.37nJ/b Adaptive-Frequency-Hopping Transceiver for Body-Area Network," Solid-State Circuits Conference, 2008. ISSCC Digest of Technical Papers, IEEE International. Feb. 2008, pp. 132–133, 208–210.
6. J.R. Bergervoet, et al., "A WiMedia-Compliant UWB Transceiver in 65 nm CMOS". Solid-State Circuits Conference, 2007. ISSCC. Digest of Technical Papers. IEEE International. 11–15 Feb. 2007, pp. 112–590.
7. T. Hollman, et al., "A 2.7-V CMOS dual-mode baseband filter for PDC and WCDMA," IEEE J. Solid-State Circuits, Jul. 2001, pp. 1148–1153.
8. A. Yoshizawa, et al., "Anti-Blocker Design Techniques for MOSFET-C Filters for Direct Conversion Receivers," IEEE J. Solid-State Circuits, Mar. 2002, pp. 357–364.
9. D. Chamla, A. Kaiser, A. Cathelin, and D. Belot, "A Gm-C Low-Pass Filter for Zero-IF Mobile Applications with a Very Wide Tuning Range," IEEE J. Solid-State Circuits, Jul. 2005, pp. 1443–1450.
10. A. M. Durham, et al., "Circuit Architectures for High-Linearity Monolithic Continuous-Time Filtering," IEEE Trans. Circuits Syst. II, Analog Digit. Signal Process, Sep. 1992, pp. 651–657.
11. S. D'Amico, et al., "A 4th-Order Active-Gm-RC Reconfigurable (UMTS/WLAN) Filter" IEEE J. Solid-State Circuits, July 2006, pp. 1630–1637.
12. S. D'Amico, V. Giannini, A. Baschirotto, "A 1.2 V–21 dBm OIP3 4th-Order Active-Gm-RC Reconfigurable (UMTS/WLAN) Filter with On-chip Tuning Designed with an Automatic Tool," in *Proc. ESSCIRC*, 2005, pp. 315–318.
13. C.-C. Hung, et al., "A Low-Voltage, Low-Power CMOS Fifth-Order Elliptic Gm-C Filter for Baseband Mobile, Wireless Communication," IEEE Trans. Circuits Syst. Video Technol., Aug. 1997, pp. 584–593.
14. S. D'Amico, M. Conta, A. Baschirotto, "A 4.1 mW 10 MHz 4th-Order Source-Follower-Based Continuous-Time Filter with 79dB-DR", IEEE Journal of Solid State Circuits, Dec. 2006, pp. 2713–2719.
15. B. Razavi, "Design of Analog CMOS Integrated Circuits". McGraw Hill, 2001. pp. 463–465.
16. K. R. Laker, W. M. C. Sansen, "Design of Analog Integrated Circuits". McGraw Hill, 1994. pp. 27–29.
17. J. Silva-Martinez, J Adut, M. Rocha-Perez, M. Robinson, and S. Rokhsaz, "A 60-mW 200 MHz Continuous-Time Seventh-Order Linear Phase Filter with On-chip Automatic Tuning System," IEEE J. Solid-State Circuits, vol. 38, no. 2, Feb. 2003, pp. 216–225.
18. M. Chen, J. Silva-Martinez, S. Roskhsaz, and M. Robinson, "A 2 Vpp, 80–200-MHz Fourth-Order Continuous-Time Linear-Phase Filter with Automatic Frequency Tuning," IEEE J. Solid-State Circuits, vol. 38, no.10, Oct. 2003,, pp. 1745–1749.

19. M.-ul-Hasan, Y. Sun, "2V 0.25 μm CMOS 250 MHz Fully-differential Seventh-order Equiripple Linear Phase LF Filter" IEEE Int. Symp. Circuits and Systems (ISCAS), 23–26 May 2005, pp. 5958–5961.
20. G. Bollati, S. Marchese, M. Demicheli, R. Castello, "An 8th-order CMOS Low-Pass Filter with 30–120 MHz tuning Range and Programmable Boost," IEEE J. Solid-State Circuits, vol. 36, no. 7, Jul. 2001, pp. 1056–1066.
21. V. Saari, M.Kaltiokallio, S. Lindfors, J. Ryynanen, K. Halonen "A 1.2 V 240 MHz CMOS Continuous-Time Low-Pass Filter for a UWB Radio Receiver" ISSCC Digest of Technical Papers, Feb. 2007, pp. 122–591.

Chapter 10
Reconfigurable Active-RC Filters with High Linearity and Low Noise for Home Networking Applications

Jan Vandenbussche, Jan Crols, and Yuichi Segawa

Abstract This paper presents a wideband reconfigurable active-RC filter designed for home networking applications. The reconfigurable filter is embedded in an analog front-end chip (AFE) for HomePNA and PLC applications. The AFE, implemented in a 0.13 μm CMOS technology, offers a high-performance analog receive and transmit path required to deliver the multi-100 Mbps targeted system level performance. With a measured SNR of 58 dB and a linearity of 95 dB, and a noise level of 3 nV/$\sqrt{}$Hz, the AFE is well suited to accommodate high data rate modulation schemes. The integrated third order filter has a programmable low-pass filter from 14–60 MHz with 1 MHz increments. A digital control loop compensated for PVT variations with 1 MHz accuracy. The AFE runs from 3.3/1.2 V supplies and consumes 152/221 mA for a full scale differential output current setting of 80 mA ptp.

10.1 Introduction

The HomePNA [1] and Power Line Communication (PLC) [2–4] standards for triple-play home networking solutions are moving to true multi-100 Mbps performance. The PLC PHY's use an OFDM modulation scheme offering currently data rates up to 200 Mbps over the electrical wiring in the house. HomePNA v3.1 is using a QAM/FDQAM modulation scheme offering data rates up to 320 Mbps with guaranteed Quality of Service (QoS). Symbol rates vary from 2 to 24 Mbaud with constellations of 2 to 10 b per symbol. The AFE is developed in 0.13 μm CMOS technology, enabling single chip integration of the home networking protocol. The current solution uses 2 chips: a digital baseband chip and the AFE. The AFE was developed for wired communication over phone line, coax, and power line, but is also capable of supporting xDSL applications.

J. Vandenbussche (✉) and J. Crols
AnSem, Leuven, Belgium
e-mail: Jan.Vandenbussche@ansem.com

Y. Segawa
Kawasaki Micro-Electronics, Makuhari, Japan

A.H.M. van Roermund et al. (eds.), *Analog Circuit Design: Smart Data Converters, Filters on Chip, Multimode Transmitters*, DOI 10.1007/978-90-481-3083-2_10,
© Springer Science+Business Media B.V. 2010

The HomePNA standard and its upcoming extensions imposed the most stringent requirements on the PHY. The frequency band spans a 4–60 MHz, making the AFE suited to operate also for mode D of HomePNA v3.1 which uses the 4–52 MHz band. Depending on which mode of the HomePNA v3.1 standard is used, output power levels (on the line) of up to 0 dBm are required. A programmable gain range of 60 dB is needed on the receive path. Compared to the previous HomePNA standard, the new HomePNA v3.1 uses coax and phone line. In order the fully exploit the coax capabilities, higher SNR numbers under high attenuation conditions are targeted. The AFE targets SNR numbers of 45 dB at 0 dB RX gain setting; resulting in a noise level requirement for the AFE of less then $3\,\mathrm{nV}/\sqrt{\mathrm{Hz}}$ and a need for above 10 b linearity. These requirements imply the need for highly linear, low noise filter and gain stages that can be reconfigured depending in the bands used. As the noise specification is tight, accurate tracking of the LPF needs to be provided as well. This paper will focus on how these requirements were handled, starting with architectural-level design considerations and circuit-level considerations for the filter and gain stages.

This paper is organized as follows. Section 10.2 explains the architecture of the AFE. Section 10.3 focuses on architectural-level design considerations and Section 10.4 describes the circuit-level considerations. Experimental results are presented in Section 10.5.

10.2 Architecture Selection

The architecture is fully optimized to deliver the most stringent specifications set by the HomePNA 3.1 standard [1]: noise levels below $3\,\mathrm{nV}/\sqrt{\mathrm{Hz}}$, above 10 b linearity, and fast gain switching (within 1% of the gain in 300 ns) are the most important specifications. The main differences between the previous HomePNA standard and the new HomePNA 3.1 are: [1] use of both phone and coax (2) extended bandwidth enabling up to 320 MBps and (3) higher output power levels up to 0 dBm. In order to fully exploit the coax with its better noise performance of $-160\,\mathrm{dBm/Hz}$, lows noise design of the AFE is key. In order to allow higher power levels tolerable on coax, linearity should be achieved even when delivering 80 mA ptp. In addition, fast switching between gain settings is required for the HomePNA standard. These three constraints have resulting in optimization both at the architectural level and the circuit level as will be explained later on. The design of the line-driver, delivering the 80 mA ptp, will not be further discussed in this paper.

Architectural-level design considerations, as will be explained in Section 10.3, resulted in the AFE architecture as shown in Fig. 10.1. The architecture is similar to ADSL transceivers [5, 6], but needed some important changes on the filter and gain stages to accommodate the stringent noise and linearity specifications. Both HomePNA and PLC operate in frequency bands in the range of a few tens of MHz. The highest frequencies are used for HomePNA, where higher order external filter networks are used to isolate the HomePNA band from others. The filters are connected

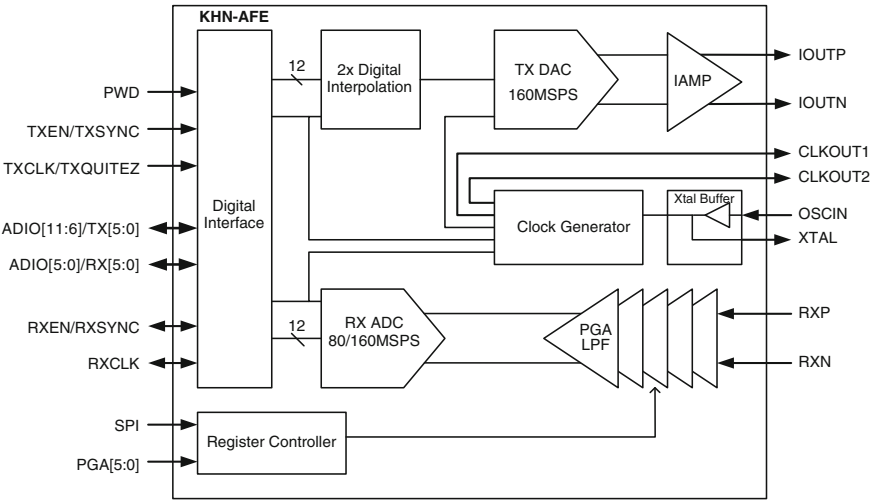


Fig. 10.1 Block diagram of the home networking AFE

to a transformer that couples the signals to the line. Signal swings of 6 Vptp, differential around the 3.3 V supply can be delivered to the transformer. Incoming signals have varying signal strength typically between 10 mV and 5 Vptp, differential. Care has to be taken to provide enough isolation between transmit and receive path because of these large output signal swings in combination with possible low input signal levels. To accommodate these various conditions both RX and TX signal path are highly programmable as will be explained next.

The transmit path comprises a classic 12 b current steering DAC [8] running from a 1.2/3.3 V supply, and a current amplifier line driver (see also Fig. 10.1). The transmit path has a programmable gain from −19.5 to 0 dB with a 0.5 dB increment. The highest gain setting delivers 80 mAptp on a 75 Ohm terminated line resulting in a voltage swing of about 6 Vptp centered around the 3.3 V supply. The current amplifier line driver is based on the regulated cascade output current mirror [7]. To accommodate the large swings above the supply, the amplifier has dedicated ESD structures and uses nMOS cascodes in a floating p-well in its output structure. The DAC has a programmable sampling rate varying between 40 and 160 MSps

The receive path comprises a programmable gain amplifier (PGA), a programmable low pass filter (LPF) and a 12 b pipelined ADC [9]. The PGA has an overall RX gain from −18 dB up to +42 dB with 1 dB increments. The LPF has been implemented as a third order Butterworth filter, the corner frequency is programmable between 14 and 60 MHz, with a 1 MHz increment. The PGA-LPF stages are implemented as active RC topologies because of the high linearity constraints. The block diagram is shown in Fig. 10.2. As the PGA stages need to switch fast between gain settings, the gain stages were separated from the filter stages, avoiding possible large transients before reaching full 10 b linearity from switching the large capacitances in the filter stages. The LPF is implemented in two stages, a

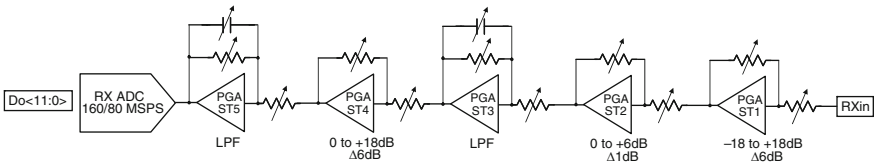


Fig. 10.2 Block diagram of the RX PGA gain and filter stages

first order section in the front of the RX chain, and a biquad section just before the ADC to avoid folding of wide band noise from the opamps in the gain stages. The LPF corner is tuned within 1 MHz.

10.3 Architectural-Level Design Considerations

10.3.1 Optimized for Noise

The full transmit and receive signal path have been optimized for noise, as this is the most stringent specification of all. For the transmit signal path the current steering DAC offers negligible noise contribution as long as the input impedance level from the IAMP is low. As the input of the IAMP is basically a diode connected NMOS, this can be easily achieved.

For the receive signal path, the architecture needed to be optimized in order to deliver the targeted $3\,\mathrm{nV}/\sqrt{\mathrm{Hz}}$. The PGA input structure is reconfigurable, such that for every gain setting (be it amplification or attenuation) the input impedance has a fixed value of $400\,\Omega$ differential, which will set the lower boundary of the noise-level. From PGA stage three on, the opamps have been down scaled to reduce power consumption.

For the data convertors a 2x oversampling ratio was chosen. A 42-taps interpolation/decimation FIR filters, resulting in a 3 dB noise reduction, is used.

A final noise source which is carefully to be considered is the jitter on the sampling clock for both DAC and ADC. The PLL is based on a self-biased active filter topology which has the ability to adjust the loop bandwidth according to the frequency. The VCO uses a four stage differential ring oscillator. Power consumption is around 35 mA for the maximum setting of 320 MHz oscillation frequency. A jitter performance of 5.3 ps rms has been simulated. The low jitter performance has been confirmed in measurements.

10.3.2 Optimized for Speed

To allow fast change in gain settings (within 1% of the gain in 300 ns), gain and filters settings have been implemented on separate PGA stages. This to avoid large settling times because of possible injected charge on the large capacitances of the

filter stages. The third order Butterworth filter is realized on stage 3 and stage 5. The gain is realized on stage 1, stage 2 and stage 4. The corner frequency of the LPF is set by switching a capacitor array. The switchable resistor array in filter stages is used for the tuning loop which tracks for corner, supply and temperature variations (see also Section 10.4).

10.3.3 Optimized for Linearity

Also for linearity both receive and transmit path were carefully mapped on a suited architecture. The HomePNA 3.1 uses 1024 QAM modulation with a PAR of 12 dB, resulting in high linearity constraints.

The main bottleneck in the linearity for the transmit side is the intrinsic linearity of the current mirror based IAMP. This was addressed by a dedicated IAMP topology, which is not further discussed in this paper.

For the receive path, the filter and gain stages needed optimization. Firstly, the gain was carefully mapped on the PGA chain, such as to optimally exploit the dynamic range of the stages. Obviously for noise considerations gain is shifted to the first stages as much as possible. To extend the dynamic range, the PGA stages and the input buffer of the ADC are running from a 3.3 V supply. Optimization of the gain mapping showed optimum results when the maximum gain on a single stage was limited to 18 dB, the maximum signal output swing in between the stages is limited to 1 Vptp, differential. The PGA stages are AC-coupled to the ADC to allow both the output of the PGA and input of the ADC buffer to operate in optimal voltage range.

Secondly, the architecture implements LPF and programmable gain such that all switches, needed to select corner frequency and gain, have been placed at a virtual ground node to achieve good linearity.

Finally, the opamps have an adaptive compensation that varies as gain settings vary. This allows the highest possible BW as the loop gain changes. This is beneficial for suppression of high frequency harmonic components. This is explained in more detail in Section 10.4. By doing so open loop GBW of 280 MHz (in typical conditions) were obtained for the maximal gain setting of 18 dB.

10.4 Circuit-Level Design Considerations

In order to accommodate noise, linearity and speed requirements, not only the architecture, but the circuit topologies themselves needed careful design. Obviously, all analog signal paths are implemented fully differential both in transmit and receive path. In addition deep n-well isolation separates transmit and receive path from the digital part which includes decimation and interpolation filtering. Apart from these generally applicable considerations, the building blocks in both the transmit path and the receive path needed dedicated solutions as will be explained next.

10.4.1 Reconfigurable Filter and Gain Stages

Both gain and LPF corner frequency needed to be programmable over a large range. This implies that every gain and filter stages is reconfigurable to offer the targeted range. Figure 10.3 shows stage 5 in the PGA stage chain, which is the biquad section realizing two of the three poles in the Butterworth filter. The LPF corner frequency is set by selecting the proper capacitor. The capacitor bank consists out of parallel branches each having a switch at the virtual ground node of the opamp. The resistor bank equally consists out of parallel branches, each branch is activated through a switch at the virtual ground. For the gain stages a similar approach has been followed. This to minimize contribution of the non-linear switch when passing large full scale signals. Because the PGA stages run from a 3.3 V supply, simple pass gates could be used without the need of gate boosting techniques.

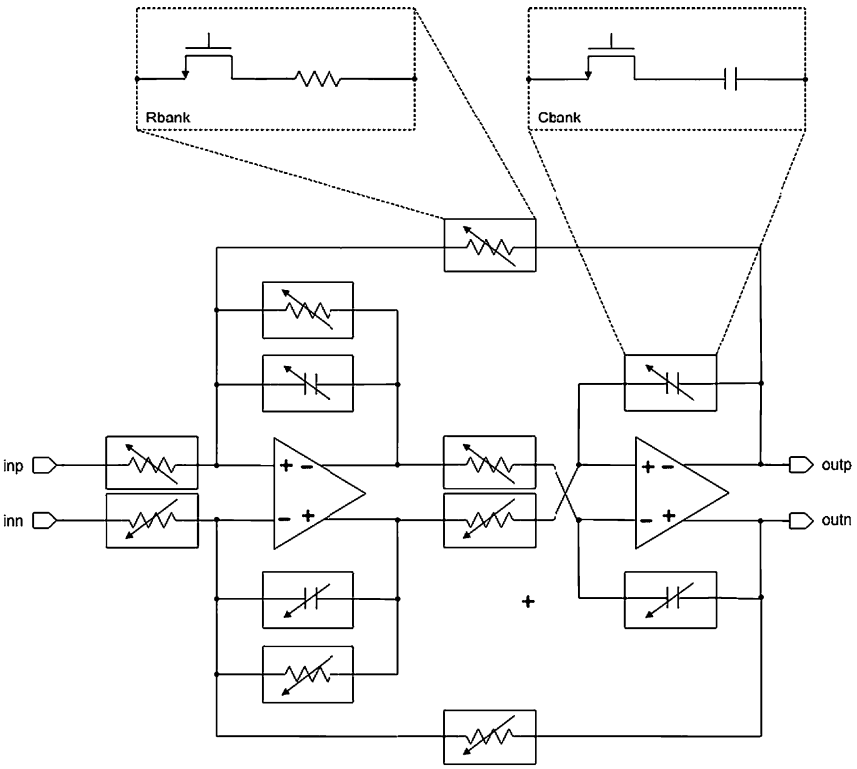


Fig. 10.3 Blockdiagram of the Biquad Tow-Thomas low pass filter section

10.4.2 PGA Opamps with Adaptive Compensation

All filter and gain stages use a miller compensated opamp as shown in Fig. 10.4. Linearity and noise are key in the design of these opamps as explained previously.

In order to achieve the targeted 10 b linearity in the stages, the gain per stage was limited to +18 dB. In addition, the compensation capacitance varies together with the gain setting. For higher gain settings less compensation capacitance is needed for stability as the feedback factor decreases. Changing the capacitance accordingly allows increased GBW for the higher gain settings, which is beneficial for suppression of distortion components at higher frequencies. The common mode feedback (CMFB) used in the opamp stages is shown in Fig. 10.5 [10]. The input stage Mn7 is a scaled down replica of the input stage Mn1 of the opamp. Because of the low impedance node vdp, and the 1/4 scaling of the input/output transistors, intrinsic stability is obtained. Using this CMFB high BW on the common-mode loop is achieved. As this CMFB structure also has an operating point in which all currents are zero, a startup circuit has been added (not shown in Fig. 10.5).

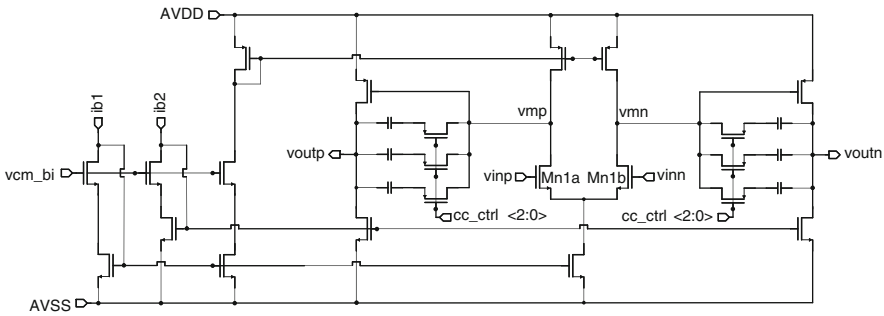


Fig. 10.4 Schematic of opamp from PGA stage 1: adaptive compensation allows to maximize BW as gain settings change

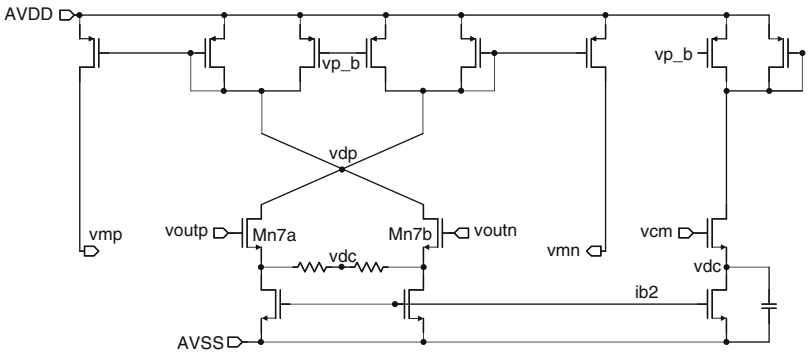


Fig. 10.5 Schematic of the common-mode feedback of the opamp: stability is guaranteed by down-scaled replica of input structure opamp

10.4.3 Tuning Loops

Because of the high programmability both in gain and filter range, tuning loops have been added.

Firstly, a DC offset tuning loop is needed to compensate for offset amplification that could saturate any of the intermediate stages in the PGA chain. At power up, or when initiated through SPI, the inputs of the RX chain are shorted, and the offset is measured through the ADC. In this calibration mode the PGA is DC coupled to the ADC. The regulator uses a 6 b current steering DAC that injects a correcting current at the output of the first stage opamp in the RX chain (nodes vmn, vmp in Fig. 10.4).

Secondly, the RC time constant of the filters is compensated for process variations, temperature and voltage drift. The strategy used is based on using a replica structure of the biquad filter section in oscillation mode [11]. The operation is as follows, see also Fig. 10.6. By removing the damping in the replica biquad section, the replica will oscillate at the corner frequency of the filter sections. The oscillation frequency is divided by 256 and compared to a reference clock of 8 MHz. The resulting clock from the oscillation replica biquad section is known in advance, it only depends on the filter corner frequency setting. This means also the amount of times the reference clock fits in the divided clock can be calculated as a function of the programmed corner frequency. These values are stored in a ROM, and the count down block is initiated with that target value from the ROM at the beginning of a tuning cycle. The deviation in number of counts after one tune cycle represents the deviation in frequency and can be used to correct the frequency. By adding and storing this deviation, the programmable code of the resistor bank in both the oscillating replica as well as the filter stages can be updated. Upon conversion of the regulating loop, the oscillating frequency and thus also the corner frequency of the filter will be within ± 1 MHz of the programmed value.

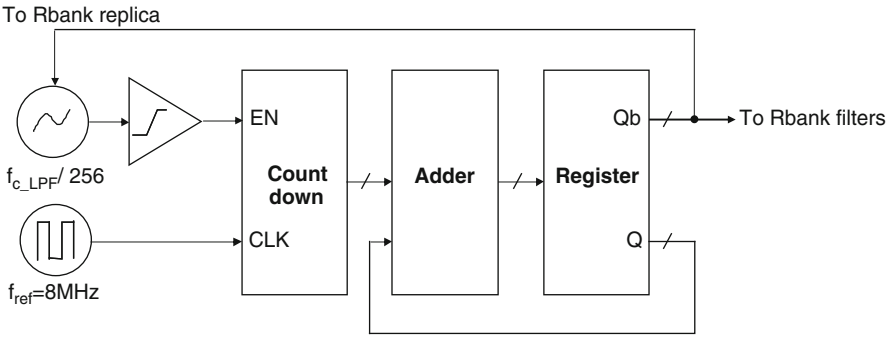


Fig. 10.6 RC tuning loop

10.5 Experimental Results

The analog front-end including programmable digital interfacing, interpolation/decimation filters and PLL have been implemented in a TSMC 0.13 μm 1P6M process with analog options for MIM and floating p-well. Only six metal layers were used. The chip area measures $2,680 \times 2,530\,\mu m$. Figure 10.7 shows a die micrograph. The current steering DAC is located at the upper right corner, it is isolated by a separate well from the remainder of the chip. The IAMP is placed next to the DAC. The outputs are carefully laid out to accommodate the peak current of up to 80 mA. Dedicated ESD structures were used to connect the output of the IAMP to the chip as the output is connected to the line via a balun, centering the voltage swing around the 3.3 V supply. ESD was designed not to add deterioration for peak values up to 5 V. The receive side is found in the lower right corner. The pipeline ADC is located at the bottom and is again shielded by isolated well. The PGA stages are placed in between the convertors at the right side. Careful layout was needed, respecting the symmetry in the opamps. The left side of the layout includes all the digital and the PLL. The digital section includes a 2x interpolation and decimation filters and all controllers for the different tuning loops as well as the programmable half/full duplex digital interface.

The chip has been assembled in a 64-QFN package and measured under various conditions. Samples are currently in production ramp-up.

Figure 10.8 summarizes the characteristics and measured performance of the AFE. Power consumption varies depending on the output power needed. For

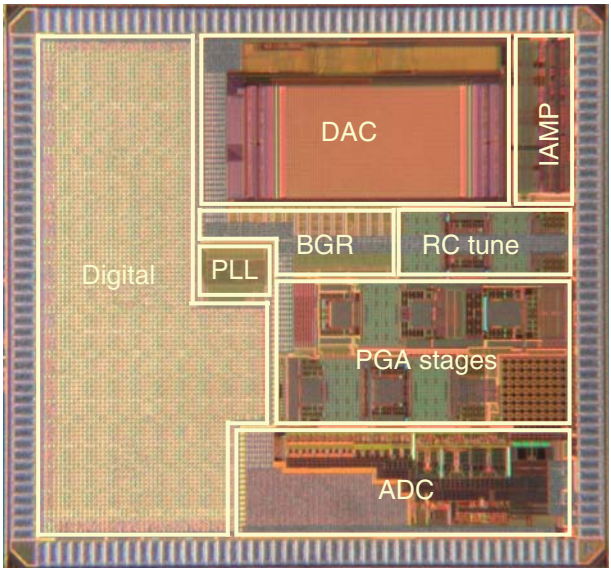


Fig. 10.7 Home networking AFE micrograph

TX path:		
DC characteristics	4-80 mA ptp	
AC characteristics:		
Fund.(Tx PGA = 0dB)	0 dBm	
SNDR (Tx PGA = 0dB)	50 dBc	
IMD (Tx PGA = 0dB)	90 dB	
Tx PGA gain	–19.5 to 0 dB with 0.5 dB increment	
Tx PGA gain accuracy	monotonic	
RX path:		
Input voltage span	8mV to 6.3 Vptp	
Input voltage noise density	< 3 nV/√Hz	
Input impedance	400 Ω	
Rx LPF f3dB	14-60 MHz	
Rx LPF pass band ripple	± 1dB	
Rx PGA gain	–18 to +42 dB with 1dB increment	
Rx PGA gain accuracy	Monotonic	
settling to 5dB PGA gain step	20 ns	
settling to 60dB PGA gain step	100 ns	
SNDR		
(Rx PGA = 42 dB, 20MHz)	9.2 bit	
(Rx PGA = –12 dB, 20MHz)	6.1 bit	
SINAD		
(Rx PGA = 42 dB, 20MHz)	38 dBFS	
(Rx PGA = –12 dB, 20MHz)	56 dBFS	
Technology	0.13µm triple-well 1P6M CMOS	
Die size (including IO)	2680x2530 µm	
Power during operation :	Rx HD [mA]	Tx HD [mA]
(HD mode, 160MSps both ADC	1.2V Digital: 35	1.2V Digital: 34
and DAC, RX gain: 42dB, TX	3.3V Digital IO: 32	3.3V Digital IO: 5
gain: 0dB)	1.2V Analog: 118	1.2V Analog: 93
	3.3V Analog: 189	3.3V Analog: 92

Fig. 10.8 Home networking AFE performance measurement summary

HomePNA 3.1, with maximum gain setting on RX/TX side and the ADC DAC running at 160 MSps, following numbers were measured in half duplex (HD) mode. The RX path consumes 35/32 mA for the digital 1.2/3.3 V supply, and 118/189 mA for the analog 1.2/3.3 V supply. The TX path consumes 34/5 mA for the digital 1.2/3.3 V supply, and 93/92 mA for the analog 1.2/3.3 V supply. For PLC applications about 20 mA less power is consumed from the analog 3.3 V supply.

All measurements were done at full clock rate of 160 MSps, unless noted otherwise.

Measurements on the DAC operating at 160 MSps show an ENOB between 12.9 and 9.7 b in the band of interest. For PLC an external line driver is used and the IAMP is bypassed. A measured PSD of the transmit path including the IAMP is shown in Fig. 10.9 for a 4 MHz single tone test. This measurement was performed for a maximum TX PGA gain setting of 0 dB delivering a differential peak-to-peak output current of 80 mA. Figure 10.10 shows two tone measurements results under the same conditions: IMD is shown vs. Tx gain. Figure 10.11 shows an MTPR measurement for a full scale IAMP current of 80 mA; the DAC is running at 160 MSps; a value of 47 dB has been measured.

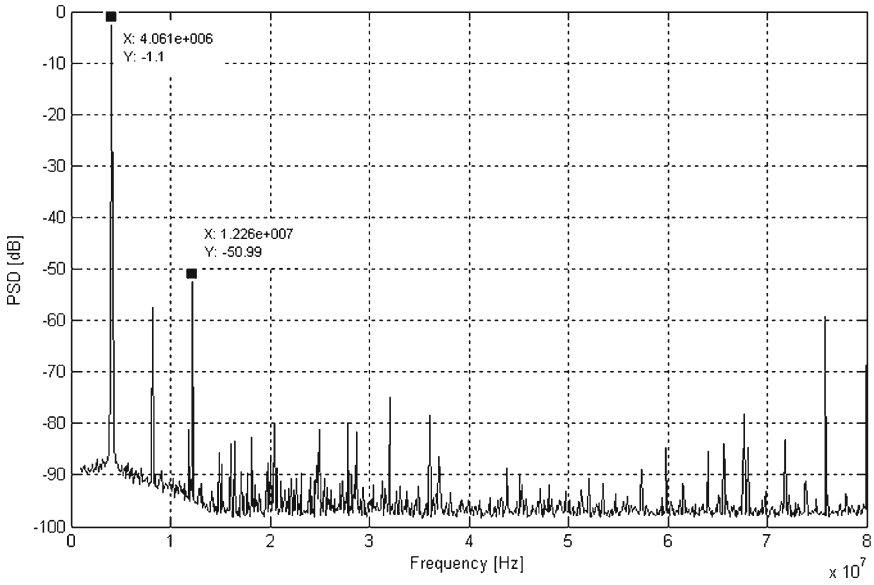


Fig. 10.9 TX power spectral density for TxPGA = 0 dB; fin = 4 MHz, internal clock @ 160 MHz, DAC @ 160 MHz

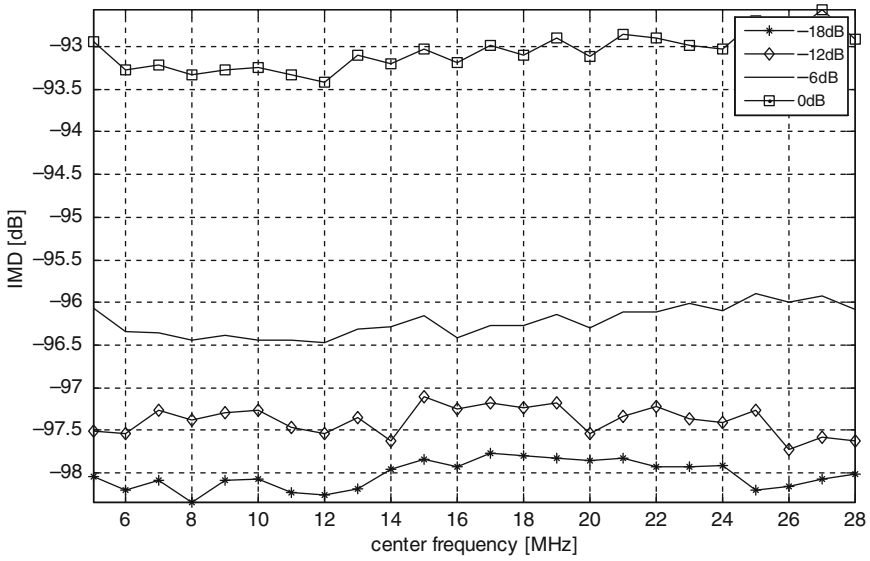


Fig. 10.10 Two-tone IMD vs. input frequency and TxPGA gain setting, DAC @ 160 MHz

Figure 10.12 shows an RX power spectral density for a 4 MHz single tone input. The ADC was run at 64 MSps, no decimation filtering was used in this measurement. Figure 10.13 shows the measured ENOB and SINAD of the RX path with the ADC running at 160 MSps.

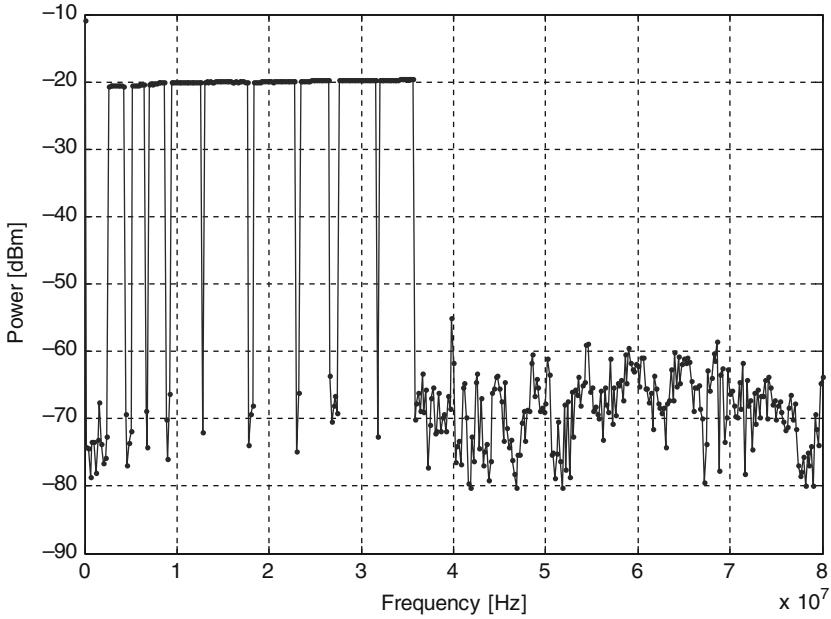


Fig. 10.11 MTPR measurement for 160 MSps operation, with maximum TX gain setting of 0 dB

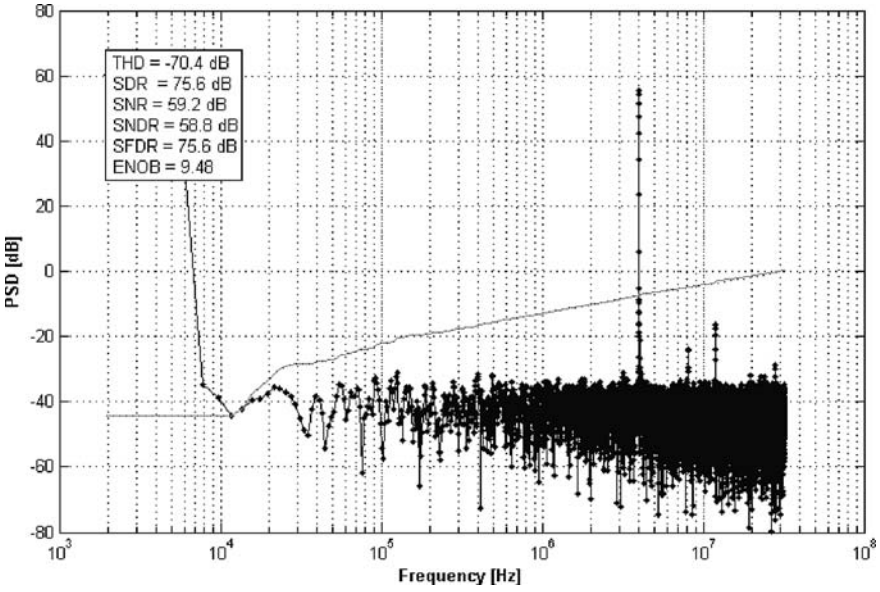


Fig. 10.12 RX power spectral density for RxPGA gain = 0 dB, −1 dBFS input at the ADC, fin = 4 MHz, internal clock@128 MHz, ADC@64 MSPS

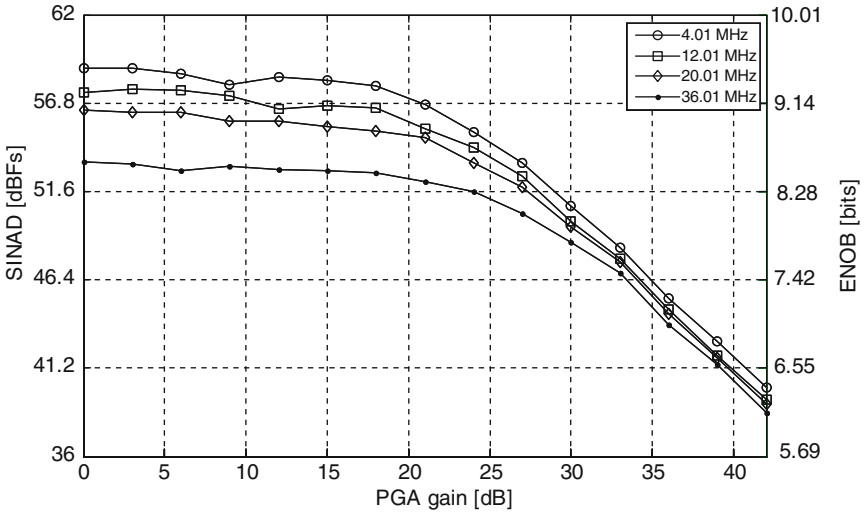


Fig. 10.13 SINAD/ENOB vs. RxPGA gain and Frequency, internal clock@160 MHz, ADC@160 MSPS, f3 dB = 37 MHz

10.6 Conclusions

A wideband reconfigurable active-RC filter designed for home networking applications has been presented. The filter is part of a front-end supporting both PLC standards and the HomePNA v3.1 standard. The design was optimized both at architectural level as well as circuit level to meet the stringent noise and linearity specifications. The architecture has been optimized to offer extended performance in terms of bandwidth and dynamic behavior of the chain compared to other front-ends that focus on DSL applications. This optimization mainly impacted the filter and gain stages design. The integrated third order filter has a programmable low-pass filter from 14–60 MHz with 1 MHz increments. A digital control loop compensated for PVT variations with 1 MHz accuracy. The chip has been implemented in a 0.13 μm 1P6M CMOS technology. The AFE has a power consumption of 152/221 mA from 3.3/1.2 V supplies, and a chip area of 2680 × 2530 μm. With a measured SNR of up to 58 dB and linearity of 95 dB, and a noise level of 3 nV/$\sqrt{}$Hz, this front-end is able to support complex modulations schemes enabling multi-100 Mbps data rates for home applications.

References

1. Home Phoneline Networking Alliance, HPNA V3.1 PHY Interface Specification, No. 06–003, May 2006.
2. HD PLC standard: http://www.hd-plc.org.

3. UPA PLC standard: http://www.upaplc.org.
4. K. Findlater et al., “A 90 nm CMOS Dual-Channel Powerline Communication AFE for HomePlug AV with a Gb Extension”, ISSCC Digest of Technical Papers, Feb. 2008, pp. 464–465.
5. Cresi, M. et al., “An ADSL Central Office Analog Front-End Integrating Actively-Terminated Line Driver”, Receiver and Filters, ISSCC Digest of Technical Papers, Feb. 2001, pp. 240–241.
6. Weinberger, H. et al., “An ADSL-RT Full-Rate Analog Front End IC with Integrated Line Driver”, IEEE J. Solid State Circuits, vol.37, no.7, Jul. 2002, pp. 857–865.
7. Serrano T. et al., “The Active-Input Regulated-Cascode Current Mirror”, IEEE Trans. Circuits and Systems, vol. 41, no.6, June 1994.
8. http://www.chipidea.com/; part no.: CI8544tl.
9. Hernes, B. et al., “A 1.2 V 220MS/s 10b Pipeline ADC Implemented in 0.13 μm Digital CMOS”, ISSCC Digest of Technical Papers, Feb. 2004, pp. 256–526.
10. Hernandez, D. et al., “Continuous-Time Common-Mode Feedback for High-Speed Switched Capacitor Networks”, IEEE J. Solid State Circuits, vol.40, No.8, Aug. 2005, pp.1610–1617.
11. Vasilopoulos A. et al., “A Low-Power Wideband Reconfigurable Integrated Active-RC Filter with 73 dN SFDR”, IEEE J. Solid State Circuits, vol.41, no.9, Sept. 2006, pp.1997–2007.

Chapter 11
On-Chip Instantaneously Companding Filters for Wireless Communications

Vaibhav Maheshwari and Wouter A. Serdijn

Abstract Instantaneous companding offers several advantages over conventional AGC techniques to deal with the high Peak-to-Average Power Ratio (PAPR) and high dynamic range (DR) associated with wireless signals in a low-voltage environment. The practical on-chip implementation of such internally non-linear systems, however, poses several challenges that arise due to process non-idealities. This paper presents the design and on-chip implementation of a companding baseband channel-select filter for WLAN 802.11a/g receivers. The filter is implemented as a fifth order Chebyshev type Switched Capacitor (SC) filter with a cut-off frequency of 10 MHz and with companding by a factor of four in IBM's 1.2 V, 130 nm CMOS technology. It achieves an almost flat Signal-to-Distortion Ratio (SDR) of around 50 dB when companding takes place in the higher end of the DR of the input signal. No AGC is required in the baseband in front of or within the filter and a reduction in power consumption by a factor of 3.3 is achieved with respect to the conventional filter designed for the same DR.

11.1 Introduction

Analog baseband filters are one of the key components used in wireless receivers for channel selection, i.e., to reject out-of-band signals before analog-to-digital (A/D) conversion. This relaxes the dynamic range (and therefore resolution) and speed requirements of the A/D converter (ADC), which would otherwise have to oversample the entire input signal containing large interferers. In a given integration technology, the dynamic range of these filters is limited by the supply voltage on the higher side and their input referred noise on the lower side. With the continuous decrease in supply voltage in integrated circuits due to the downscaling of modern digital CMOS

V. Maheshwari (✉) and W.A. Serdijn
Electronics Research Laboratory, Faculty of Electrical Engineering, Mathematics and Computer Science, Delft University of Technology, Mekelweg 4, 2628 CD, Delft, The Netherlands
e-mail: {v.maheshwari,w.a.serdijn}@tudelft.nl

A.H.M. van Roermund et al. (eds.), *Analog Circuit Design: Smart Data Converters, Filters on Chip, Multimode Transmitters*, DOI 10.1007/978-90-481-3083-2_11,
© Springer Science+Business Media B.V. 2010

technologies, in order to keep the high dynamic range of the filters, one needs to decrease their noise figure. But every 3 dB decrease in the noise power costs double the capacitor area used in the chip and around double the power consumption. An alternate, less expensive way to accommodate the high dynamic range of the input signal within the low dynamic range of the filters is to use an Automatic Gain Control (AGC) in front of the filter. However, since most wireless receivers are subject to the presence of interferers during minimum desired signal reception, the allowable AGC gain in front of the filter is limited. Therefore, often the AGC operation is distributed throughout the filter, further amplifying the signal as interferers are attenuated [1]. In digital communication systems, the AGC settings of the receiver are set during the preamble or the midamble of the data frame. Once the gain settings are set, they are not changed for the rest of the frame even though the input signal power might vary, e.g., due to the high Peak-to-Average-Power Ratio (PAPR) of the signal or due to variations in the transmission channel. This mandates extra headroom in the dynamic range of the filter, which leads to higher power consumption. Also, lowering the supply voltage will lead to an even higher AGC resolution if we want to keep the noise figure of the filter unchanged. This may, in turn, lead to a longer settling time of the AGC loop, which in many applications is not acceptable. Besides, extra headroom in the dynamic range of the filter still needs to be provided because of input signal variations.

Companding[1] offers several advantages over conventional AGC techniques [2]. In principle, companding can be considered a type of AGC in which the gain control works at all times during the data transmission. Companding systems generally include an input gain element, a signal processor and an output gain element. The input gain element compresses the high dynamic range input signal, which is then processed by the low dynamic range signal processor (the filter in our case) followed by expansion using the output gain element. In order for the output of the signal processor not to be disturbed by dynamic modifications of the gain at the input end, one must control the systems state variables accordingly [3]. As an example, Fig. 11.1 shows the block diagram of a companding lossy integrator (lossless for $a = 0$, where $a \leq 0$ is a fixed feedback gain factor) [2]. Here, the state variable $w(t)$ is defined as

$$w(t) = g(t) \cdot x(t) \tag{11.1}$$

where $g(t)$ is a function of $w(t)$ and $x(t)$ is the output of the integrator without companding, i.e. when $g(t) = 1$. The time varying gain $g(t)$ is used to achieve compression at the output of each integrator, thereby reducing its dynamic range requirements and power consumption. Since integrators are the main building blocks and power horses of analog filters, companding thus helps in reducing the power consumption of the filter. From a practical on-chip implementation point of view, however, companding, being an ELIN system [2], poses several challenges that arise due to process non-idealities. It is often difficult to ensure the accuracy and timing of compression and expansion (g and $1/g$ in Fig. 11.1) due to mismatch

[1] Companding is a portmanteau of compressing and expanding.

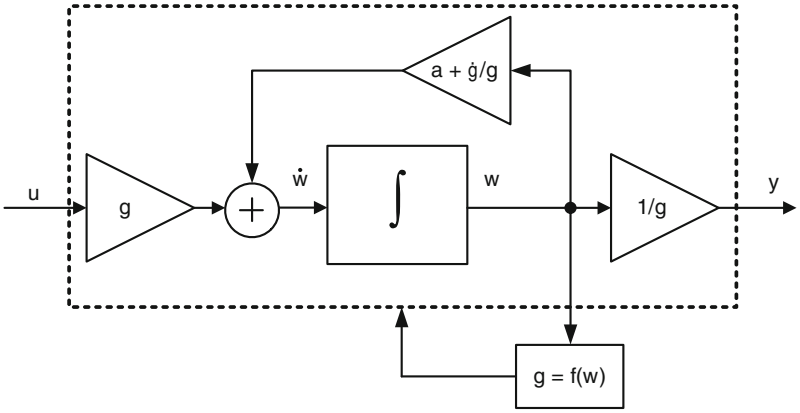


Fig. 11.1 Companding lossy integrator

and process variations in a practical design, which gives rise to distortion [2]. Successful designs have been made using companding in log domain circuits to achieve very high dynamic range filters [4] but circuits in the log domain are mainly intended for very low power, e.g., biomedical applications. Companding in the discrete time domain [2] using Switched Capacitor (SC) circuits [5], on the other hand, offers several advantages. The compression and expansion functions are carried out using an array of switched capacitors. The capacitor mismatch in CMOS technologies is normally less than 0.1%, thus making it possible to use them to compress and expand the signal with high accuracy. The discrete time implementation allows the control block to have sufficient time to measure the signal strength and transmit the appropriate digital signals to execute companding, in time before the next signal sample is ready to be processed. In this paper, we present a companding baseband SC filter with a fifth-order Chebyschev type frequency response, 10 MHz cut-off frequency, 100 MHz clock frequency and companding by a factor of four designed for WLAN applications. In the next section, the design and the operation of companding SC filter is explained in detail.

11.2 Companding Switched Capacitor Filter Implementation

A practical implementation of companding filters can be achieved by using a piecewise-constant gain function g [5]. In this case, the $\dot{g}/g$ term in the feedback block of the integrator shown in Fig. 11.1 can be implemented by state variable updating given by the relation [2]:

$$w\left(t_k^+\right) = w\left(t_k^-\right) \cdot \frac{g_k}{g_{k-1}} \tag{11.2}$$

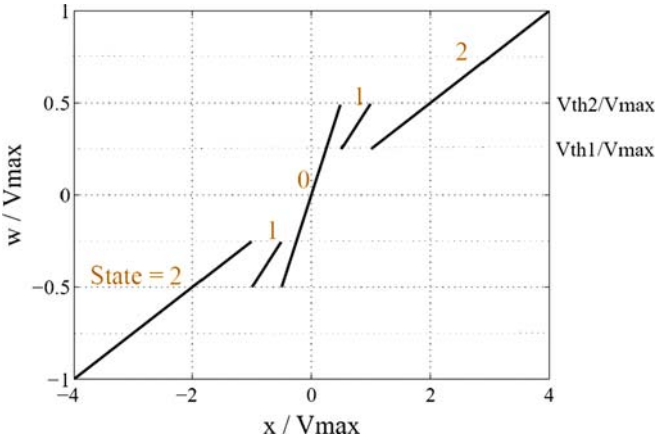


Fig. 11.2 Mapping from x to w

In the above equation, k is used as an index to represent different values of g that appear in time, t_k is the time instant at which the value of g changes and $w(t_k^-)$ and $w(t_k^+)$ denote the limit of $w(t)$ as time t approaches t_k from left and from right respectively.

Figure 11.2 shows an example of x to w mapping, in which the consecutive values of g (1, 1/2, 1/4) differ by a factor of 2 [5,6]. V_{max} is the maximum allowable voltage that could saturate the output of the opamps. We define three states 0, 1 and 2 by the variable *State* corresponding to the values 1, 1/2 and 1/4 of g respectively. In such a case, the updating of state variable w amounts to either doubling it or halving it whenever there is a change in g. Comparators are employed to detect the crossings of w through the thresholds (V_{th1} and V_{th2} in Fig. 11.2, where $V_{th1} < V_{th2}/2$ to avoid instability) and change g accordingly.

A companding filter using a piecewise-constant function g can be easily implemented using SC integrators. Figure 11.3a shows the parasitic insensitive SC discrete integrator in which C_{s1} and C_{s2} denote the value of the input and output sampling capacitors, respectively, and C_I denotes the value of the integration capacitor. Let $\Phi 1$ and $\Phi 2$ be the two non-overlapping clock phases. The small de-glitching capacitor C_{dg} does not play a role in the signal charge redistribution. It is used to prevent glitches in the opamp's output by providing negative feedback during the brief intervals when the non-overlapping clock phases are both low, and the feedback path of the opamp is otherwise open-circuited [7].

The input signal is sampled in phase $\Phi 1$ during which the output of the opamp remains constant. During phase $\Phi 2$, the charge from the sampling capacitor C_{s1} is transferred to the integrating capacitor C_I. Figure 11.3b shows the companding discrete integrator along with state variable update circuitry used to double and half the output voltage depending on whether g increases or decreases by a factor of two respectively [5, 6]. An array of capacitors is used at the input and the output for gain scaling. The integrator in Fig. 11.3b is shown as used in the first stage of the

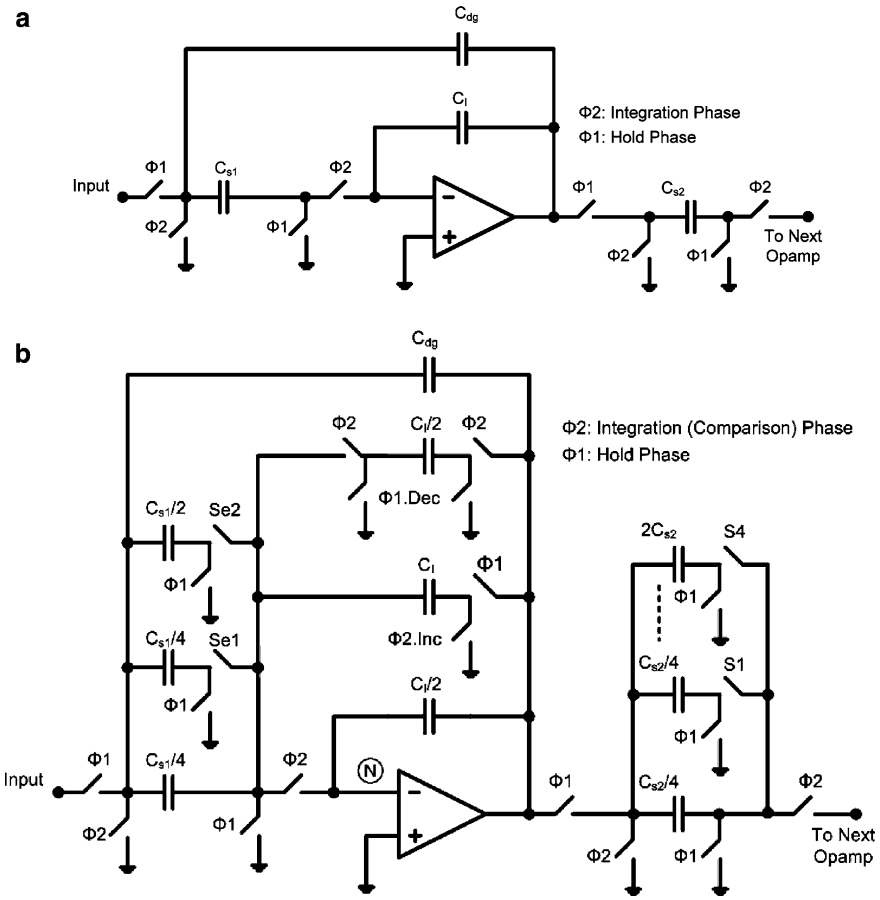


Fig. 11.3 (**a**) SC discrete integrator, (**b**) Companding SC discrete integrator

filter in which the capacitor array at the input is used to compress the input signal ($g = 1$, 1/2 or 1/4) with the help of control signals S_{e1} and S_{e2} derived from the output of the first stage. Similarly, a capacitor array is used at the output of the last stage of the filter for expansion ($1/g = 1$, 2 or 4). For the intermediate stages, the capacitor array at the output of one stage is combined with the array at the input of the following stage to form a single array, which is controlled using the logic signals S1, S2, S3 and S4 derived from both the stages. In this way, the expansion factor of one stage is combined with the compression factor of the following stage to implement the equivalent inter-stage gain. The control signals *Inc* and *Dec* are used to double or halve the opamp's output in order to update the integrator's memory whenever the input gain is increased or decreased by a factor of two, respectively.

Figure 11.4 shows the timing diagram of the control signals with respect to the two non-overlapping clock phases Φ1 (hold phase) and Φ2 (integration phase). The comparison of the opamp's output voltage is done in Φ2 and appropriate control

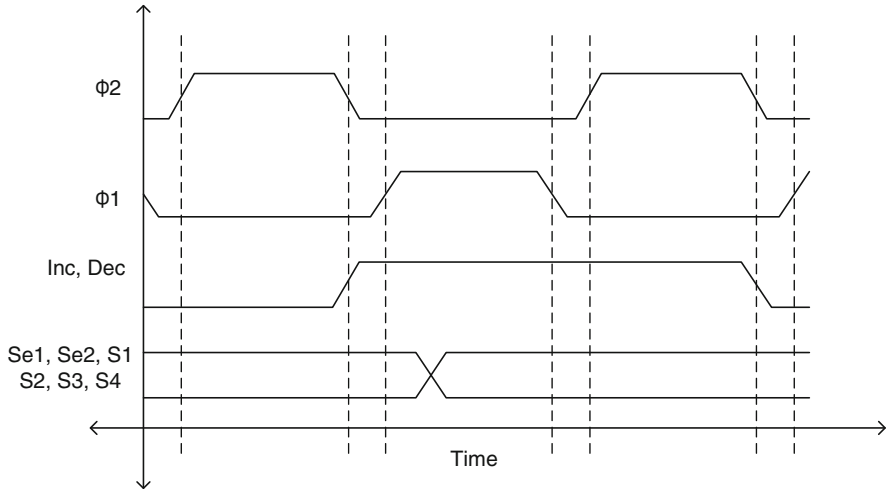


Fig. 11.4 Timing diagram of the control signals

signals are released during Φ1 to be ready to perform companding in the next integration phase. In this way, the control circuits need not be designed to be fast and there is sufficient time for the signals to propagate in the chip before the next integration phase starts. However, there is one exception; control signal *Dec* is used during phase Φ1 itself to halve the voltage across the integration capacitor by discharging one half of it as shown in Fig. 11.3b. Since the capacitor is discharged to a common mode voltage, speed is not important and the capacitor can be discharged accurately during phase Φ1.

Note that in Fig. 11.2, threshold V_{th2} is lower than V_{max} that could saturate the output of the opamps. The choice of these thresholds should be made carefully. In SC filters, when a signal with high amplitude and at a frequency close to the filter cut-off frequency is applied to the input, the jumps between the output voltages of the opamps in consecutive cycles of Φ2 can be very high, especially in *State* 0 when there is no compression. For example, the output voltage of the opamp can be less than but close to V_{th2} in one cycle of Φ2. In the next cycle of Φ2, it can jump to a much larger voltage before a decision to compress the signal is made. Therefore, threshold V_{th2} should be made lower than V_{max} to accommodate such a signal step. Although the signal is allowed to go beyond threshold V_{th2} in the final *State* 2 as shown in Fig. 11.2, part of the dynamic range may be compromised because noise puts a lower limit to the minimum value of V_{th2} that can be used. Since the filter noise is fixed based on the minimum required sensitivity, it should be ensured that V_{th2} is high enough such that compression caused by either the interferer or the desired signal does not bring down the desired signal to below signal sensitivity levels. The lower limit of V_{th2} is further set by process non-idealities like opamp's DC offset. As explained in the next section, DC offset gives rise to even order distortion. For a given value of DC offset, a lower V_{th2} will result in lower Signal-to-Distortion ratio since companding starts early when the signal power is

low. Thus, the minimum value of V_{th2} may put a limit on the maximum amplitude of the signal that the companding filter can accommodate without saturating any of the opamps.

An alternative solution could be to increase the over sampling ratio of the SC filter so that the signal is sampled faster and thus, the voltage jumps would be smaller. In such a case, V_{th2} can be set to a higher value. Also, having a higher over sampling ratio eases the anti-aliasing requirements of the pre-filter. However, a faster clock increases the power consumption and also makes the design of switches challenging to achieve low distortion, especially at high input frequencies. A higher clock frequency can also make the value of the sampling capacitors (C_{s1} and C_{s2} in Fig. 11.3a) very small making it unrealistic to realize the array of capacitors in Fig. 11.3b on chip. Higher thresholds may also demand higher slew rate of the opamp's output stage. So, there are several parameters: over sampling ratio, threshold voltages, noise, opamp's DC offset, distortion and slew rate that should be considered in order to optimize the system for best performance.

We can estimate how much power that can be saved using companding by a factor of four for an increase in dynamic range of 12 dB. A ladder type low pass filter has both feedforward and feedback paths between consecutive stages as shown in Fig. 11.5. The figure shows an example of one of the intermediate stages of the

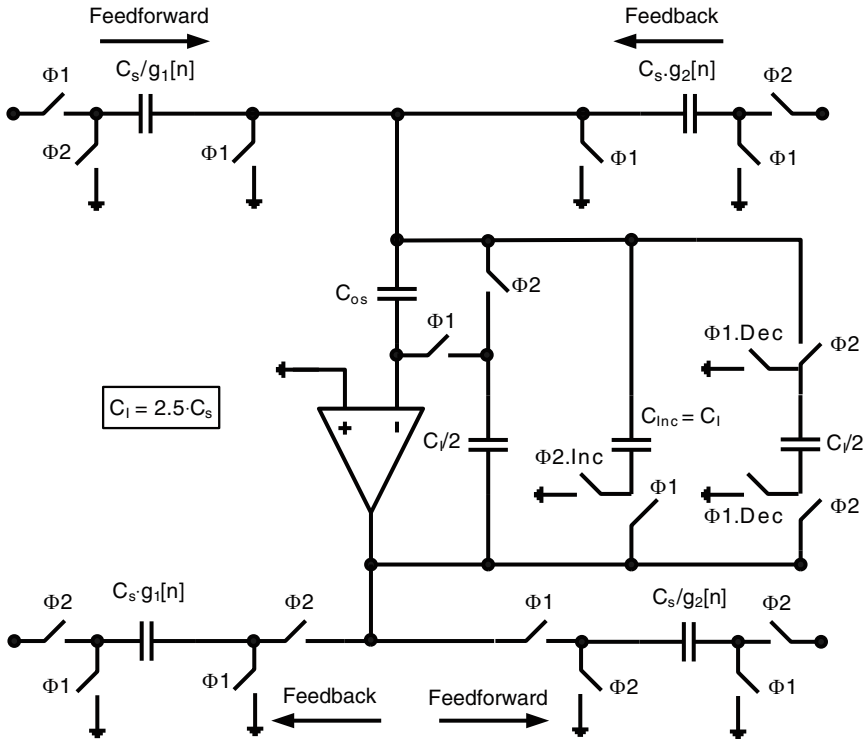


Fig. 11.5 Filter stage using companding SC discrete integrator

filter implemented using the companding discrete SC integrator. For simplicity, we assume that all sampling capacitors have a value C_s and the integration capacitor has a value C_I. C_{os} is the opamp's DC offset storing capacitor to cancel the DC offset in integration phase Φ2 using Correlated Double Sampling (CDS). In Fig. 11.5, $g_1[n]$ and $g_2[n]$ can take a value from the set (1/4, 1/2, 1, 2, 4). The first and the last stage of the filter are slightly different but the approach and the results are similar. For a Chebyshev type, fifth order low pass filter, C_I is approximately 2.5 times C_s. Using these values, it can be estimated how much power overhead is needed to implement instantaneous companding. As an example, in Fig. 11.5, the feedback factor β_c and the load C_{Lc} in both clock phases Φ1 and Φ2 are given by:

$$\beta_c(\phi 1) = \frac{C_I/2}{C_I/2 + C_{os}} \tag{11.3}$$

$$\beta_c(\phi 2) = \frac{C_I}{C_I + C_s/g_1[n] + C_s \cdot g_2[n]} \tag{11.4}$$

$$C_{Lc}(\phi 1) = C_I + C_s/g_2[n] + \beta_c(\phi 1) \cdot C_{os} \tag{11.5}$$

$$C_{Lc}(\phi 2) = C_s \cdot g_1[n] + \beta_c(\phi 2) \cdot (C_s/g_1[n] + C_s \cdot g_2[n]) \tag{11.6}$$

The capacitive load C_{Lc} during phase Φ1 dominates because of the presence of extra capacitance C_I. The feedback factor β of the opamp in negative feedback is kept the same during both clock phases and for different gain settings with the help of dummy capacitors so that the opamps are stable and have the same step response under all conditions. It follows:

$$\beta_c(\phi 1) = \beta_c(\phi 2) = \beta_c \tag{11.7}$$

$$C_{os} = (C_s/g_1[n] + C_s \cdot g_2[n])/2 \tag{11.8}$$

For the non-companding case, when the filter is designed for the same input referred noise as with companding, β_{nc} and C_{Lnc} are given by

$$\beta_{nc} = \frac{C_I}{C_I + 2 \cdot C_s} \tag{11.9}$$

$$C_{Lnc} = C_s + \beta_{nc} \cdot 2 \cdot C_s \tag{11.10}$$

A 2-stage miller compensated opamp is used in the filter in order to achieve high DC gain and to handle large output voltage swing. Let the power consumed by the first stage and second stage of the opamp be *P1* and *P2*, respectively, for the non-companding case. Note that if β decreases by a factor x and C_L increases by a factor y then the current in the first and second stages of the opamp have to be increased by a factor of x^2 and y, respectively, in order to maintain the same noise and settling behavior of the opamp. From above equations, using the worst case values of $g_1[n]$ and $g_2[n]$, $x^2 \approx 5.4$ and $y \approx 3.5$. Since *P2* is much greater than *P1* due to the high capacitive load, the power overhead for each opamp turns out to be close

to but less than four times. The opamp in the expansion stage consumes as much power as other opamps in the non-companding case ($\approx P2$). Therefore, the power overhead becomes 4.2 times. Ideally, a 12 dB improvement in the dynamic range of the filter should result in 16 times reduction in power consumption. Therefore, it is estimated that instantaneous companding results in approximately 3.8 times power savings for a 12 dB improvement in the dynamic range as compared to a conventional filter. There will be extra power consumed by the comparators used in each stage of the filter and it has not been taken into account yet. This will be discussed in a later section.

11.3 Opamp's DC Offset Cancellation

A companding SC filter is an Externally Linear Internally Non-linear (ELIN) system and any spurious signal arising from within the system would be affected by the non-linearity present in the system. In our case, the unwanted signal is the opamp's DC offset, which gives rise to even-order distortion. Even-order distortion can be explained by the fact that, for a sinusoidal input, the DC offset has the same sign in both positive and negative half cycle of the sine wave. The mechanism of appearance of even-order distortion from the opamp's DC offset can be explained as follows. Referring to the companding SC integrator shown in Fig. 11.3b, the opamp's DC offset V_{OS} affects the operation of companding in two ways. By applying charge conservation principle on the virtual ground node N of the opamp, it can be shown that every time the voltage across C_I is halved or doubled, there is an injection of charge of value $C_I \cdot V_{OS}$ into the integration capacitor, adding a voltage V_{OS} to the output signal. The second mechanism happens due to the presence of $k \cdot V_{OS}$ in the output voltage V_o of the integrator when it is connected in negative feedback. Here, k is a proportionality constant, which depends on the negative feedback factor. Let the voltage stored across the integration capacitor (i.e., the desired output voltage) be V_{od}, then we have $V_o = V_{od} + k \cdot V_{OS}$. When compression happens by a factor of two, V_{od} gets halved but the V_{OS} component of V_o remains the same. Finally, when the signal is expanded by two, we get $V_o = V_{od} + 2kV_{OS}$. Thus, for a sinusoidal input, the expanded signal would be the result of the addition of V_{od} and a rectangular pulse similar to the one as shown in Fig. 11.6. The error signal has an amplitude $k \cdot V_{OS}$, has a frequency twice the input frequency and a pulse width equal to the time interval during which the companding takes place. Both of the above described mechanisms give rise to even order distortion, but from simulations it was found that the second one is the most dominant.

In an SC ladder type filter, the output of the opamp in each stage is sampled in both phases, once for the feedforward path and once for the feedback path as shown in Fig. 11.5. Thus, the DC offset of the opamps should be eliminated in both phases. The Correlated Double Sampling technique [8] eliminates the DC offset in the integration phase only. In order to get rid of the DC offset in the feedforward path as well, one can use a non-inverting delay element (e.g., an offset compensated flip

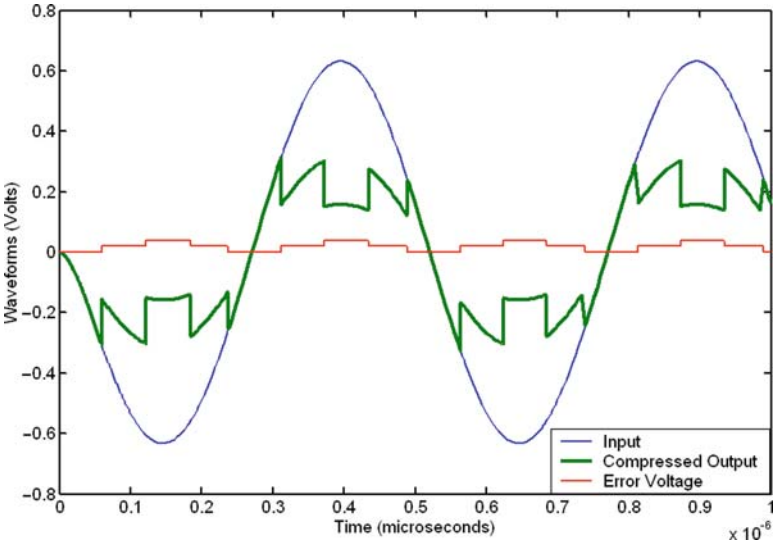


Fig. 11.6 Waveforms to illustrate the effect of the opamp's DC offset in a companding switched capacitor filter

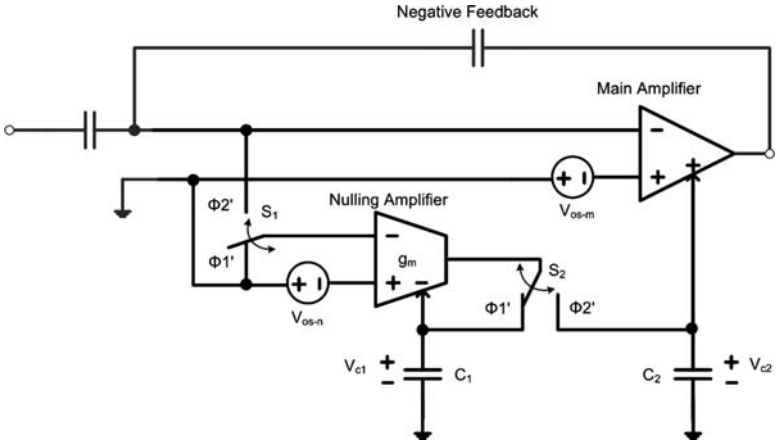


Fig. 11.7 Continuous-time AZ amplifier using feedforward technique

around Track and Hold amplifier [8]) that can sample the offset free output in the integration phase and make it available, offset free, in the hold phase for the next stage. Since this delay element works in almost unity gain feedback configuration, there will be an overhead of 1.25 times more power. An alternative solution, which is cheap in terms of power, is to use a continuous-time Auto-Zeroed (AZ) amplifier using feedforward technique [8] (Fig. 11.7). However, it has some practical limitations. As discussed next, in this design, this technique gives a residual offset of

500 μV under worst case process and mismatch conditions. Therefore, we use both CDS as well as a continuous-time AZ amplifier to achieve the desired THD of less than −50 dBc.

Figure 11.7 shows a continuous-time AZ amplifier using a feedforward technique [8]. The basic principle behind this circuit is to use a low offset auxiliary amplifier to cancel the offset of the main amplifier connected in negative feedback. The circuit operates in two non-overlapping clock phases Φ1′ and Φ2′. During phase Φ1′, the nulling amplifier is auto zeroed and its nulling voltage V_{c1} is stored on the capacitor C_1 at the end of phase Φ1′ and held during phase Φ2′. This offset free amplifier is then available in phase Φ2′ to sense the DC offset of the main amplifier at its input and generate a nulling voltage V_{c2} on capacitor C_2 to cancel the DC offset of the main amplifier. This voltage is held constant during the next phase, Φ1′. Theoretically, the final residual DC offset of the main amplifier is given by the total offset of both amplifiers divided by the low frequency DC gain of the nulling amplifier. This is based on the assumption that the DC offset of both the amplifiers referred to their auxiliary input port is the same as their input referred DC offset respectively [8]. For instance, if the DC offset of each amplifier is 10 mV and the DC gain of the nulling amplifier is 60 dB then it should result in a residual DC offset of 20 μV, which is sufficiently low for the companding filter. But in reality, some charge sharing happens when the switch S2 switches back and forth between C_1 and C_2 and thus creates a jump in the nulling voltages. This jump in voltage depends on the size of the capacitors C_1 and C_2 with respect to the parasitic capacitances in the nulling amplifier that are involved in the charge sharing. It also depends on the magnitude and sign of the nulling voltages V_{c1} and V_{c2}, the worst case being when they have opposite signs.

The design of the nulling amplifier mainly consists of 3 parameters – the unity-gain bandwidth (UGBW) of the nulling amplifier, the value of the nulling voltage storing capacitors C_1 and C_2 and the clock frequency. The residual offset occurring due to the charge sharing can be reduced by either increasing the UGBW, by increasing the value of C_1 and C_2 or by decreasing the clock frequency while keeping other parameters fixed. C_1 and C_2 are chosen reasonably to be integrated on chip. The nulling amplifier creates a pole-zero pair at its UGBW in the overall transfer function of the main amplifier and is thus limited by the settling behavior. Finally, the clock frequency is limited by the anti-aliasing requirements. In this design, we chose the auto zeroing clock frequency at 12.5 MHz for WLAN 802.11 g (at 10 MHz for WLAN 802.11a), at which there is no signal present. Any spurious signal is further attenuated by the low pass filtering of the nulling amplifier. The capacitor values of C_1 and C_2 are 1 and 2 pF, respectively. The UGBW is 1 MHz so as not to affect the settling. This design results in a worst case residual offset of 500 μV. The nulling amplifier consumes only 40 μW compared to 6 mW consumed by the main amplifier.

Besides the opamp's DC offset, any mismatch coming from capacitors would give rise to odd-order distortion. However, the 3σ value of the MiM capacitor mismatch for the lowest capacitor value used in the design (100 fF) is less than 0.3% and thus is less of a problem for a target THD of −50 dBc for WLAN applications.

11.4 WLAN Receiver Baseband Signal Chain

An important practical limitation for companding filters to be used in wireless communication systems is the presence of large interferers in the input signal. Since the noise in a companding filter is dependent on the signal level, a large interferer can trigger companding and cause the noise to rise, which can corrupt the Signal-to-Noise Ratio (SNR) of a simultaneously present small desired signal. Thus, it is important to make sure that the interferers do not affect companding otherwise some linear pre-filtering should be used before the companding filter.

For a WLAN receiver, a direct down conversion architecture is assumed. Figure 11.8 shows how signal levels vary throughout the receiver chain for a range of desired signal strength and worst case adjacent and alternate-adjacent interferers. The receiver consists of an antenna, a band pass filter, a low-noise amplifier, mixers, AGC loop for RF front-end, low pass channel-select filters, AGC loop for baseband and ADCs for the I and Q channels. The RF portions of the receiver are assumed to have the following gains (as used in Fig. 11.8): −3 dB for the (off-chip) band pass filter (BPF), 0 or +20 dB (selectable) for the low-noise amplifier (LNA) and +10 dB for the mixers [9]. The entire receiver has a noise figure of 5 dB and the RF front-end is assumed to have a noise figure of 4 dB. After down conversion, an amplifier provides a gain of 10 dB, and is followed by a first order anti-aliasing

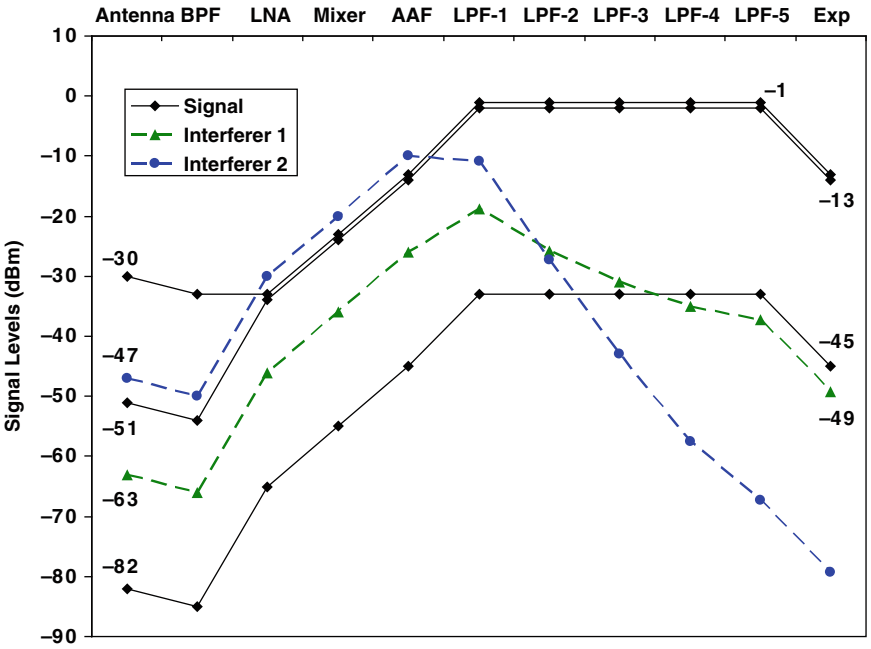


Fig. 11.8 WLAN 802.11a receiver gain distribution plot for 6 Mb/s rate. Interferer 1 is at 20 MHz and Interferer 2 is at 40 MHz

filter (AAF) with a noise figure of 18 dB. A PAPR of 10 dB is assumed and extra headroom of 6 dB is given. A fifth order, Chebyshev type, companding SC low pass filter is used for channel-selection. It has a gain of 12 dB in the first stage (LPF-1). An extra amplifier (Exp) is used to expand (recover) the signal at the output, which has a gain of −12 dB. Clipping of signal at any stage in the filter (LPF-1 through LPF-5) is avoided due to companding. The entire filter has 0 dB gain and 30 dB noise figure. The rest of the baseband consists of another AGC loop with voltage gain amplifiers (VGA), to provide fine tuning in the gain settings, and ADCs.

Note that the interferers at the first stage of the filter are low enough not to trigger companding. Also, it is assumed that some of the channel-select filtering for the 20 MHz interferer is provided using digital filtering in the DSP to attain the required SNR.

11.5 Simulation Results

The baseband filter for WLAN 802.11a/g receivers is implemented as a fifth-order, 0.1 dB in-band ripple Chebyshev low pass ladder filter (LPF) using companding SC integrators in IBM's 130 nm, 1.2 V CMOS process. The cut-off frequency is fixed at 10 MHz and the sampling clock frequency equals 100 MHz.

The filter is implemented in a differential structure using 2-stage miller compensated opamps with a low-frequency loop gain of 60 dB, loop gain-bandwidth product of 350 MHz and slew rate of 300 $V/\mu s$. The switches are implemented as CMOS transmission gates using voltage boosted clocks.

Figure 11.9 shows the plots of Signal-to-Distortion Ratio (SDR) vs. the input signal power in dBV_{rms} for a single-tone and a two-tone test. In the case of single-tone test, the input frequency is 2 MHz and the Signal-to-Total Harmonic Distortion Ratio is plotted in Fig. 11.9 for both cases, with and without companding. The even-order distortion component of the SDR in the companding case is also shown. It can be observed that both odd-order and even-order distortion behave similarly and are almost flat during companding. For the two-tone test, the input frequencies are 3 and 4 MHz and the Signal-to-Inter-Modulation Distortion (IMD) Ratio is plotted. These plots are obtained for the worst case even-order distortion due to the residual offset of 500 μV in each opamp. It can be observed that after companding starts, the SDR becomes almost flat around 50 dB and companding by a factor of 4 results in an improvement of 12 dB in dynamic range. In Fig. 11.8, we can see the presence of adjacent and alternate adjacent interferers along with the in-band signal all along the five stages of the filter. These interferers produce inter-modulation distortion components, which fall in-band with the desired signal. It can be expected that companding should also affect these inter-modulation products. Simulations show that companding does degrade the inter-modulation distortion components as compared to non-companding case. However, the distortion components are still too low to affect the SDR plot shown in Fig. 11.9 any further. Figure 11.10a and b show the

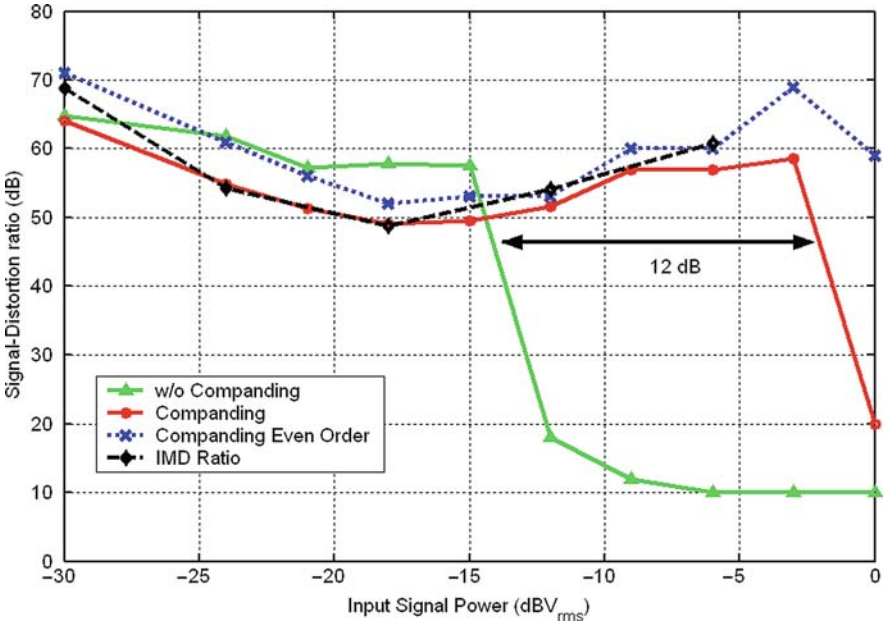


Fig. 11.9 Signal-to-Distortion Ratio vs. input signal without companding (*green line*), with companding (*red line*), with companding, even order only (*blue dotted line*) and the signal-to-intermodulation distortion ratio resulting from a two-tone (3 and 4 MHz) test (*black dotted line*)

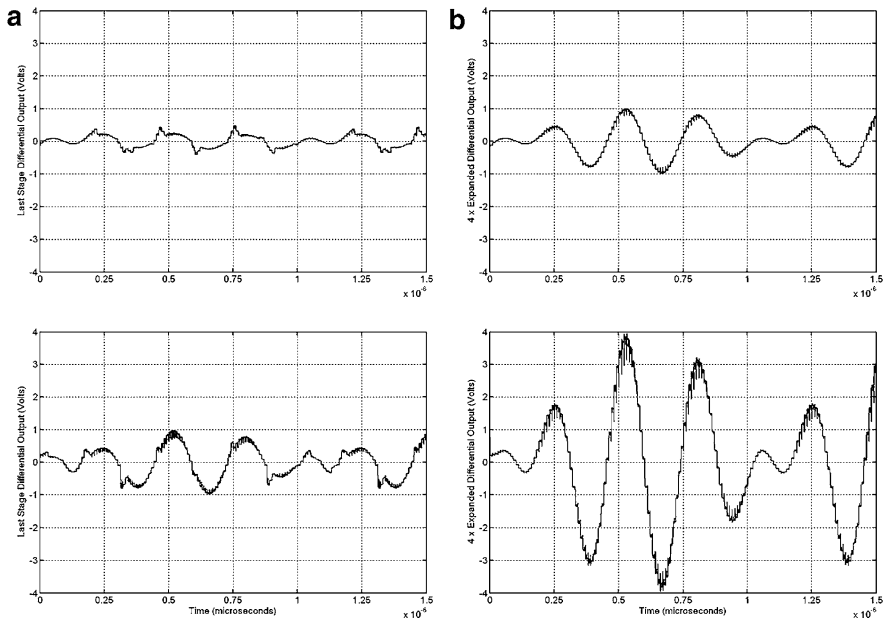


Fig. 11.10 Two-tone test voltage waveforms at the output of the (**a**) last stage of the companding filter and, (**b**) after expansion for two different input signal power strength $-18\,\mathrm{dBV_{rms}}$ and $-6\,\mathrm{dBV_{rms}}$

differential compressed output of the last stage of the filter and after expansion, respectively, for the two-tone test, for two values of input signal power $-18\,\mathrm{dBV_{rms}}$ and $-6\,\mathrm{dBV_{rms}}$.

It was estimated that companding by a factor of four should result in reduction in power consumption of the SC filter by four times for a given dynamic range. However, the control circuitry, the comparators and the output expansion amplifier consume extra power and reduce the power savings achieved by companding. The total power consumed by the filter is 31.3 mA (38 mW) of which the overhead is 5.3 mA (6. 4 mW). Therefore, this companding filter consumes 3.3 times less power than a conventional filter designed for the same dynamic range.

11.6 Summary

With the continuous decrease in supply voltage in integrated circuits due to the downscaling of modern digital CMOS technologies, the SNR of analog baseband filters is limited by the signal swing. In order to design low power channel select filters, it is important to reduce the large input dynamic range associated with wireless signals to close to the minimum required Signal-to-Noise and Distortion Ratio (SNDR). Automatic Gain Control (AGC) has been traditionally used to reduce the input dynamic range of wireless signals but it is only limited to slow variations in the input signal. Due to high Peak-to-Average Power Ratio (PAPR) and sudden channel variations, extra headroom of around 12 to 16 dB needs to be given, which results in wastage of power.

Instantaneously companding baseband filter using SC circuits is presented as a solution to deal with the PAPR. Although SC circuits, as compared to their continuous-time counterparts, make it easier to implement companding filters, since companding filter is an ELIN system, any spurious interference arising from within the filter gives rise to distortion. One of the major problems is the opamp's DC offset, which gives rise to even order distortion. In this design, both Correlated Double Sampling (CDS) and continuous-time auto-zeroing is used to cancel the DC offset to below the maximum acceptable level for all process corners. For the same reason, a good layout of the design is very crucial for good performance.

An instantaneously companding, fifth-order, baseband channel-select low pass SC filter is implemented in IBM's 130 nm, 1.2 V CMOS process for WLAN applications. For an extra gain of 12 dB in dynamic range, the filter consumes 3.3 times less power than a conventional filter. Process and mismatch variations are taken into account in simulations. Results show that an almost flat SDR of around 50 dB is achieved when companding takes place.

Acknowledgments The authors would like to thank Prof. Yannis P. Tsividis for his support and many important discussions on the design of the companding filter, and Frank van der Goes and Sandeep Mallya for important inputs in circuit design and layout.

References

1. M. T. Ozgun, Y. Tsividis and G. Burra, "Dynamic power optimization of active filters with application to zero-IF receivers," IEEE J. Solid-State Circuits, vol. 41, pp. 1344–1352, Jun. 2006.
2. Y. Tsividis, "Externally linear, time-invariant systems and their applications to companding signal processors," IEEE Trans. Circuits and Syst. II, vol. 44, pp. 65–85, Feb. 1997.
3. E. Blumenkrantz, "The analog floating point technique," Proc. IEEE Symp. Low Power Electron, vol. 1, pp. 1549–1550, 1995.
4. M. Punzenberger and Christian C. Enz, "A 1.2-V Low-Power BiCMOS Class AB Log-Domain Filter," IEEE J. Solid-State Circuits, vol. 32, pp. 1968–1978, Dec. 1997.
5. N. Krishnapura, Y. Tsividis, K. Nagaraj and K. Suyama, "Companding switched capacitor filters," Proc. IEEE Symp. Circuits and Syst., vol. 1, pp. 480–483, Jun. 1998.
6. V. Maheshwari, Wouter A. Serdijn and John R. Long, "Companding baseband switched capacitor filters and ADCs for WLAN applications," IEEE Symp. Circuits and Systems, ISCAS 2007, pp. 749–752, May 2007.
7. H. Matsumoto and K. Watanabe, "Spike-free SC circuits," Electron. Lett., vol. 8, pp. 428–429, 1987.
8. Christian C. Enz and Gabor C. Temes, "Circuit techniques for reducing the effects of op-Amp imperfections: autozeroing, correlated double sampling, and chopper stabilization," IEEE Proceedings, vol. 84, pp. 1584–1614, Nov. 1996.
9. O. Jeon, Robert M. Fox and Brent A. Myers, "Analog AGC circuitry for a CMOS WLAN receiver," IEEE J. Solid-State Circuits, vol. 41, pp. 2291–2300, Oct. 2006.

Chapter 12
BAW-IC CO-Integration Tunable Filters at GHz Frequencies

Andreia Cathelin, Stéphane Razafimandimby, and Andreas Kaiser

Abstract This paper presents a particular type of GHz frequencies high-quality Silicon integrated filters using BAW resonators. By enhancing BAW resonators with active Silicon "intelligence", process and temperature variations of such high quality factor band-pass filters may be compensated. After presenting some theoretical aspects, this paper presents the design of a frequency tunable BAW filter together with the implementation of its tuning circuitry. System-in-Package (SiP) co-integration aspects between Silicon and BAW technologies are also presented.

12.1 Introduction

BAW devices are piezoelectric resonators working in the frequency range from 1 to 10 GHz and typically they show quality factors of about 1,000. One of the potential advantages of such a technology is the compatibility of the BAW process with standard silicon processing technology and environment, thus above IC co-integration may be foreseen in some cases. For almost 10 years now, several players in the scientific academia and industry have been using such devices to build on-chip monolithic high quality factor filters and oscillating systems at GHz frequencies. Nevertheless, process dispersions on the thickness of the physical layers composing the piezoelectric device lead to a shift of the BAW resonator characteristic frequencies and thus a shift of the filter's center frequency. Moreover, BAW resonators suffer from a thermal drift of around 20 ppm/°C. Therefore, the need of designing electronically tunable BAW filters and adjacent automatic tuning circuitry appears for SoC/SiP integration in order to correct process and/or temperature deviations. The tunability of such systems may offer also interesting features for systems re-configuration. This paper is composed as follows: Section 12.2 gives a short insight on the BAW

A. Cathelin (✉) and S. Razafimandimby
STMicroelectronics, 850 rue Jean Monnet, 38926 Crolles Cedex, Grenoble, France
e-mail: andreia.cathelin@st.com

A. Kaiser
IEMN – ISEN, 41 bd Vauban, Lille, France

A.H.M. van Roermund et al. (eds.), *Analog Circuit Design: Smart Data Converters, Filters on Chip, Multimode Transmitters*, DOI 10.1007/978-90-481-3083-2_12,
© Springer Science+Business Media B.V. 2010

technology and in Section 12.3 the types of filters that may be built with such resonators are presented. In Section 12.4 we will discuss about the concepts developed for *tunable* BAW resonators. Section 12.5 presents a practical design case for a W-CDMA post LNA filter. Section 12.6 discusses about tuning circuitry implementation for such filters and Section 12.7 presents a practical design case. Finally, Section 12.8 will permit us to conclude this paper.

12.2 BAW Technology

12.2.1 BAW Resonators

BAW resonators are typically composed of three parts: the electrodes, a piezoelectric layer and an isolation part. The isolation is obtained with an air gap for TFBAR (Thin Film Bulk Acoustic Resonator) and with a Bragg reflector for SMR (Solidly Mounted Resonator) (see Fig. 12.1). The principle of the isolation part is that a change in the impedance affects the amount of acoustic energy that is reflected and transmitted. Creating a discontinuity at material boundaries allows breaking the transmission of an acoustic wave in the materials. Hence, Bragg mirror consists of several pairs of alternatively high and low acoustic impedance $\lambda/4$ material layers. Thus, most of the signal is confined in the piezoelectric material and not transmitted to the substrate. The other layers of the BAW resonators structure also influence the resonator characteristics. In particular the plate electrodes introduce a capacitor C_o in parallel with the mechanical resonator, as well as mechanical loading of the resonator, thus reducing the resonance frequency.

A more detailed cross-section view of a SMR BAW resonator is given in Fig. 12.2. Typically, the piezoelectric material used is the Aluminum Nitride (AlN), while for inter-layer compatibility and metal resistivity Molybdenum (Mo) electrodes are employed. The Bragg reflector is obtained with a multiple stack of SiN and SiOC materials. A mechanical loading layer may be employed or exploited on the top of the top electrode in order to slightly change (lower) the resonance frequencies of a resonator.

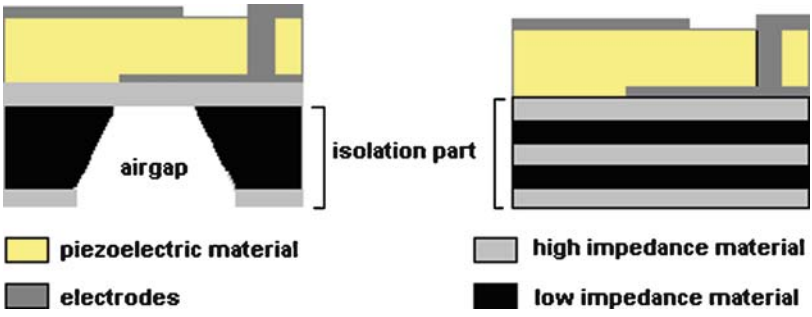


Fig. 12.1 Cross-section of different types of BAW resonators (*left side: TFBAR, right side: SMR*)

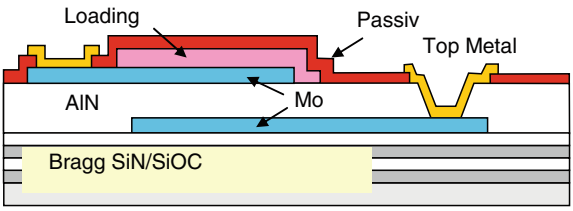


Fig. 12.2 Detailed cross-section of an SMR BAW resonator (courtesy to European Commission IST 027003 Mobilis project)

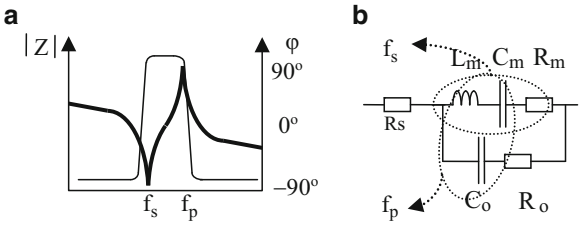


Fig. 12.3 (**a**) BAW resonator impedance. (**b**) M-BVD electrical model

12.2.2 Electromechanical and Electrical Model of a BAW Resonator

Two models are currently used: Mason and BVD model (Butterworth Van Dycke). The first one is a 1D modeling taking into account the mechanical load of the different layers used in the BAW process by acoustic and electromechanical equations. It translates mechanical forces into electrical variables. The second one represents the BAW resonator's electrical behavior (see Fig. 12.3a) by a network of lumped components (see Fig. 12.3b). The BAW resonator is characterized by a series resonance (f_s) and a parallel resonance (f_p) also called anti-resonant frequency. It is equivalent to very low impedance at f_s and to high impedance at f_p. Out of $f_s - f_p$ band, it is seen as a capacitor value C_o (see Fig. 12.3a). Moreover, the different characteristic elements of a BAW resonator are closely linked and tuning one of them directly is impossible.

The following equations permit to characterize a BAW resonator (one may notice that they are the same as any other piezoelectric resonator, such as quartz).

$$f_s = \frac{1}{2\pi} \cdot \frac{1}{\sqrt{L_m C_m}} \tag{12.1}$$

$$f_p = f_s \cdot \sqrt{1 + \frac{C_m}{C_o}} \tag{12.2}$$

$$Q_s = \frac{1}{\omega_s R_m C_m} \cdot \frac{R_o}{R_o + R_s} \tag{12.3}$$

$$Q_p = \frac{1}{\omega_p R_m C_m} \cdot \frac{C_o}{C_o + C_m} \tag{12.4}$$

$$k_t^2 = \frac{\pi^2}{8} \cdot \frac{C_m}{C_o} \tag{12.5}$$

$$Z(\omega) = \frac{1}{j\omega C_o} \cdot \frac{1 + \frac{j\omega}{Q_s \omega_s} - \left(\frac{\omega}{\omega_s}\right)^2}{1 + \frac{j\omega}{Q_p \omega_p} - \left(\frac{\omega}{\omega_p}\right)^2} \tag{12.6}$$

Where: f_s is the series or resonant frequency with its associated quality factor Q_s; f_p is the parallel or anti-resonant frequency with its associated quality factor Q_p and ${k_t}^2$ is the electromechanical coupling factor, which gives a measure of the "spacing" between the two resonant frequencies. Typical values for these parameters, for the materials described in Fig. 12.2, are f_s and f_p around 2 GHz, Q_s and Q_p around 1,000 and ${k_t}^2$ of about 6%.

The module of the resonator impedance $|Z(\omega)|$ is determined by the resonator area, which is reflected by the C_o term in Eq. 12.6. Typical values that can be implemented on-chip provide impedances from 30 to $1,000\ \Omega$.

12.3 BAW Resonator Filters

Two main categories of filters exist [1, 2]: ladder structures (see Fig. 12.4a) and lattice structures (see Fig. 12.4b). Ladder filters are single-input/single-output filters with two notch frequencies while the lattice ones are fully differential. Starting from an elementary resonator (R_s) as series resonator in the direct path, a second resonator type called R_p located in the shunt path of the filter is obtained by loading the resonator by an extra oxide layer. This extra layer shifts the characteristic frequencies thanks to the loading effect, thus lowering the resonance frequencies. In this paper, R_p will be annotated by a dot.

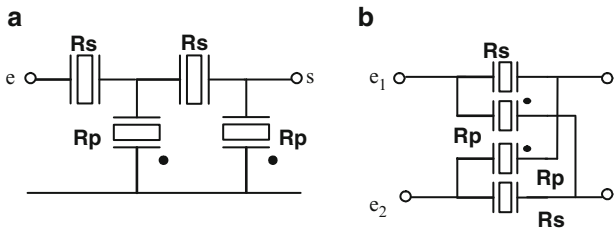


Fig. 12.4 Resonator filters: (**a**) two-stage ladder (**b**) one-stage lattice

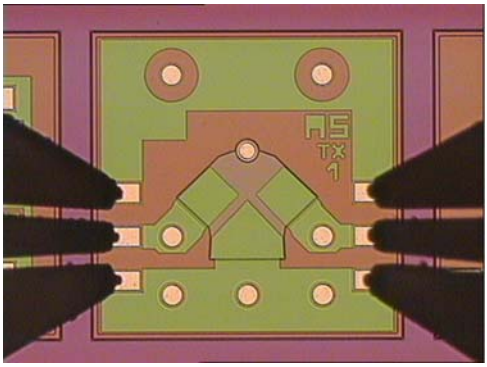


Fig. 12.5 Photomicrograph of an SMR BAW ladder filter (application related design) (courtesy to European Commission IST 027003 Mobilis project)

It is interesting to notice that the BAW technology permits the monolithic integration of such filters on one chip, thus all the filter designs are fully matched to their application, the electrodes are designed with the best suited shape and all the interconnects are minimized. For example, a photomicrograph of a ladder filter for a specific application is given in Fig. 12.5.

12.3.1 BAW Ladder Filters

As BAW resonators behave like a short circuit at f_s and like an open circuit at f_p, the series frequency of R_s is aligned to the parallel frequency of R_p in order to create a pass-band filter function. In this way, a loading material is added on the top of R_p to reduce its resonant frequencies as illustrated on the SMR of the Fig. 12.6. In the case of a ladder filter, f_s of R_p and f_p of R_s create the notch frequencies. In fact, R_p at f_s forms a RF path to ground and R_s at f_p cuts the RF signal transmission. If R_s and R_p have the same size, that is the same C_o (same impedance module), each cascaded stage brings 6 dB attenuation out of the band (see Fig. 12.6).

12.3.2 BAW Lattice Filters

The lattice filter can be analyzed in the same way as the ladder one. No notch frequency is produced because of the existence of a perpetual RF path. Lattice filter behavior can be explained by analyzing the transfer function of such a structure (see Fig. 12.7a). Indeed, the filter transfer function is defined by Eq. 12.7:

$$H = \frac{Z_p - Z_s}{Z_p + Z_s} = \frac{\left|\frac{Z_p}{Z_s}\right| e^{j\Delta\varphi} - 1}{\left|\frac{Z_p}{Z_s}\right| e^{j\Delta\varphi} + 1} \tag{12.7}$$

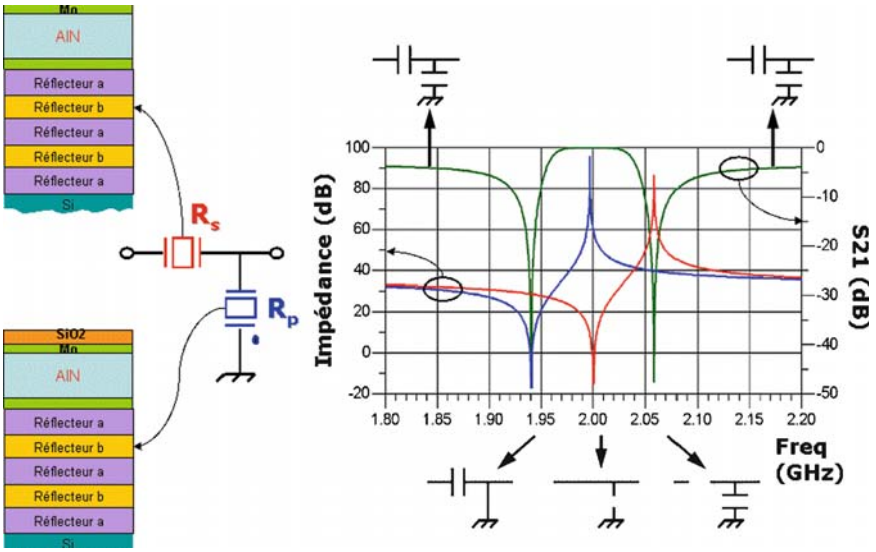


Fig. 12.6 BAW ladder filter principle

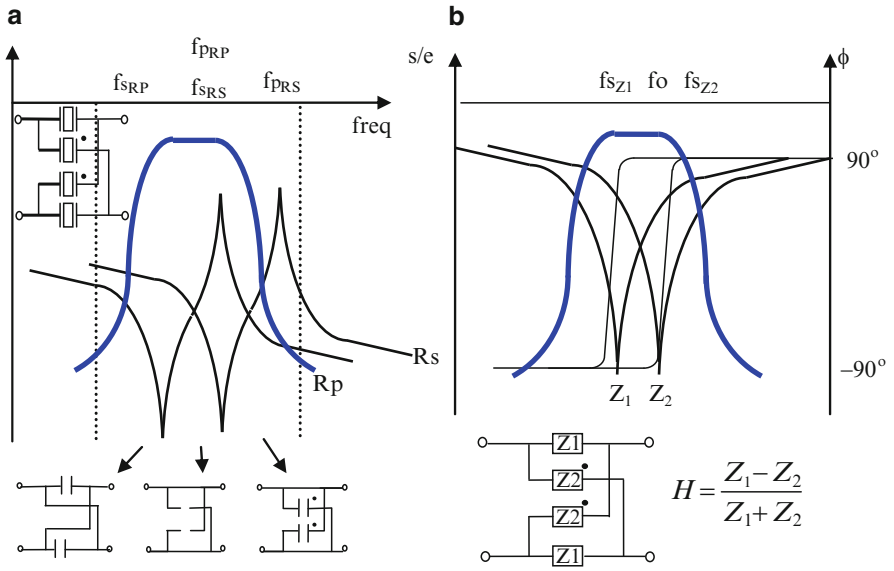


Fig. 12.7 BAW lattice filter principle: (**a**) typical approach, (**b**) phase constructive phenomenon

When series and parallel arm impedances are equal and their phases are opposite, we obtain an optimal condition to transmit the RF signal (see Fig. 12.7a). Contrary to that, when the impedances of the different branches are equal in magnitude and phase, a high attenuation is obtained.

In fact, this new way of understanding lattice filters is based on a phase constructive phenomenon (see Fig. 12.7b – only the resonant frequency of the series and parallel resonators are drawn together with the resulting filter transfer function). Indeed, in this kind of filters, only the low impedance resonant frequency f_s is used in order to build the filter transfer function. The anti-resonance frequency is of no use in the filter transfer function.

To tune this type of filters, we thus have to tune the resonant frequency of each resonant structure unless one of the resonant frequencies can be made no more useful. Notice that if a common control quantity between R_s and R_p exists, this can facilitate the filter tuning.

Finally, lattice filters exhibit a better selectivity than ladder filters. For the same given filter mask, lattice structures employ less BAW resonators. For example, for mobile communication standards, the adjacent channels attenuation is much more significant when using lattice filters. Furthermore, differential structures eliminate the constraints on even-order non-linearity.

For all these considerations, the work presented in this paper will rely on lattice filters.

12.3.3 BAW Filters Synthesis Method

The technique used to synthesize a BAW filter is based on a classical design technique for polynomial LC-filters in which the coupling concept is used.

In order to realize a BAW filter, the first step is to define or to choose a synthesis method that allows the introduction of the BAW resonator electrical model. A brief state of the art on classical filter synthesis exhibits two theories: the image parameters theory and the effective parameters theory. Only the second one turns out to be useful in our case. While the starting point of the first method is the effective attenuation (directly linked to the insertion losses) which characterizes the behavior of the network, the effective parameters authorize more freedom in the architecture which is a valuable feature in our case given the nature of the architecture to be realized. The transfer function of the chosen filter can be defined by several categories of functions among which the most common are Butterworth, Tchebychev, generalized Tchebychev.

The full development of this BAW filter synthesis theory is out of the scope of this paper, nevertheless it has been largely described in [3].

If the goal of the implementation is to obtain *tunable* BAW filters, then some passive (and if possible tunable) elements should exist in the vicinity of the BAW devices. Taking into consideration all these remarks and the pre-cited filter synthesis theory, the principle schematic of a resultant lattice tunable BAW filter is given in Fig. 12.8.

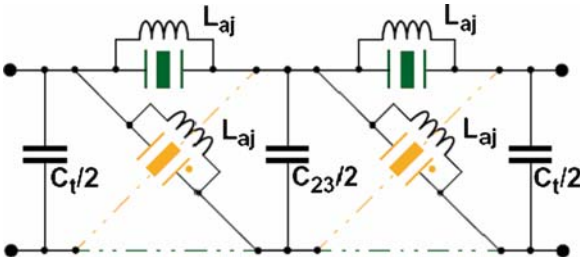


Fig. 12.8 BAW lattice filter with tuning potentiality

12.4 Tunable BAW Resonators

The goal of the exercise is to find an electrical cell containing at least one BAW resonator, which has the electrical behavior of one single BAW resonator, but which is electrically tunable over frequency. As it has been shown in Eq. 12.7 the transfer function of a lattice filter is depending on the impedances of the series and parallel resonators. If we use *tunable* resonators this means that their impedance is tunable and thus provides frequency *tunable* filters.

The goal is to correct a filter's process and temperature dispersions in the frame of the given filter mask. In classical use of ladder and lattice filters, both series and parallel frequencies of BAW resonator have to be shifted by the same ratio in order to shift the filter shape properly without any changes.

A series capacitor will increase the series resonant frequency in the theoretical limit of the anti-resonance frequency that remains constant (see Fig. 12.9a). A parallel one will decrease the anti-resonant frequency while it makes the series frequency unchanged (see Fig. 12.9b).

A tuning component with an opposite phase allows reaching the opposite phenomena. A series inductor reduces the series frequency (see Fig. 12.9c) whereas a parallel one will increase the parallel frequency (see Fig. 12.9d). Nevertheless, inductors create additional resonances by interacting with the C_o capacitance of the BAW resonator.

Integrated variable capacitors (varactors) in advanced IC processes have only a limited tuning range. Controlling both resonant frequency f_s and anti-resonant frequency f_p with only variable capacitances is unachievable on a large band.

Nevertheless, as seen before, the use of lattice filters is preferred. Indeed, the use of the BAW resonator resonant frequency (f_s) by eliminating the anti-resonant frequency offers a better opportunity to make lattice filter tunable. For this purpose, a parallel inductance will be used in order to resonate with C_o at f_p. Thus, only the resonant frequency (that is f_s) will be exploited. Using this inductance permits to push away f_p rather than eliminate it which is nevertheless sufficient.

In order to be able to tune separately the two characteristic frequencies of a resonator, we associate to a single BAW resonator an inductance in parallel and then a series capacitor (see Fig. 12.10, [4]). A tunable value for the inductance permits a

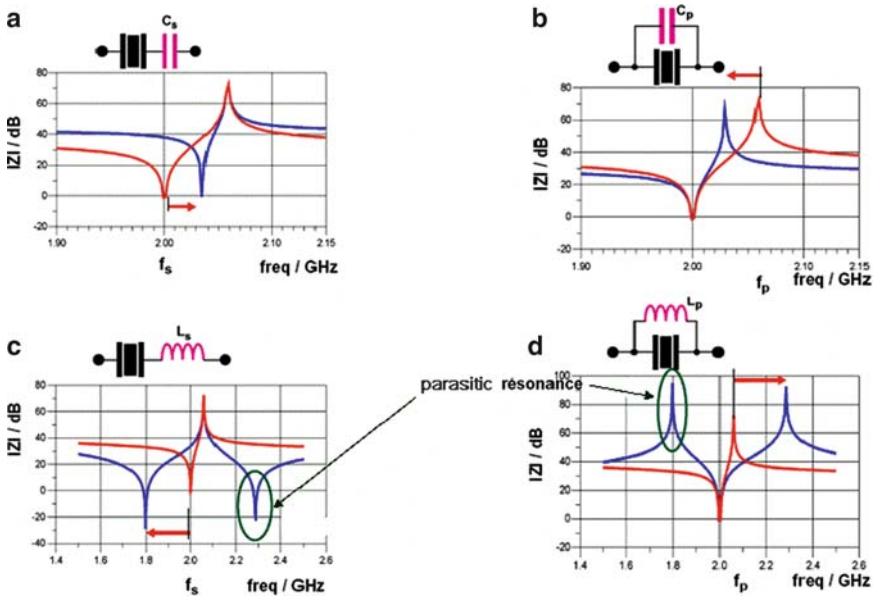


Fig. 12.9 Tuning a BAW resonator (**a**) with a series capacitance, (**b**) with a parallel capacitance, (**c**) with a series inductor and (**d**) with a parallel one

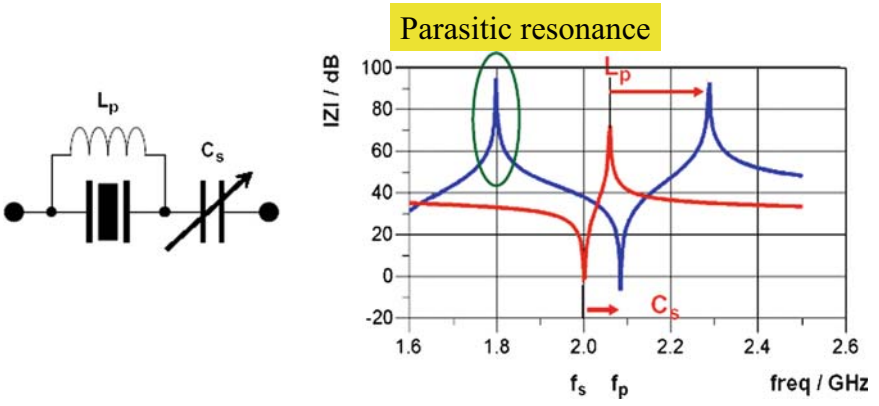


Fig. 12.10 Tunable BAW resonator cell and its impedance variation

large tuning value for the anti-resonant frequency, while a tunable capacitor permits (within its own variation range) the variation of the resonant frequency towards the theoretical limit given by the new value of the parallel frequency.

Other tuning techniques have also been developed, for example by using a negative capacitor circuitry instead of the inductor in Fig. 12.10 [5].

12.5 Design of an Electronically Tunable BAW Filter for Zero IF W-CDMA Receivers

This section presents a practical design case, where we focus on the implementation of a tunable BAW filter to be placed in a zero-IF reception chain for W-CDMA applications. This filter, typically a SAW filter in most of commercial hand-sets, used to exist in the reception chain between the LNA and the down conversion mixers and is aimed to block any undesired signals coupling from the transmit path. The goal is to correct filter's process and temperature dispersions in the frame of a given filter mask. Figure 12.11 gives the frequency mask specification for this filter [6].

The general filter structure is derived from the theory very shortly presented in Section 12.3 (see also Fig. 12.8) and also by using the BAW tunable cell presented in Fig. 12.10. This filter structure is depicted din Fig. 12.12.

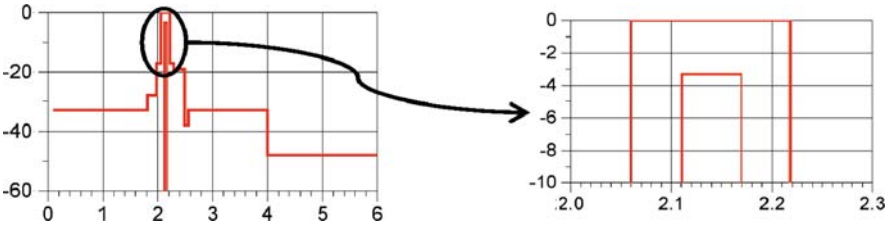


Fig. 12.11 Post-LNA filter in a zero-IF Receiver; filtering mask

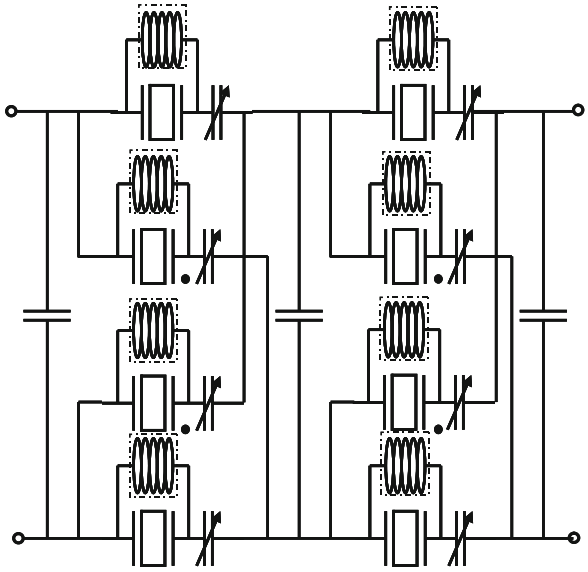


Fig. 12.12 Electrically tunable BAW filter

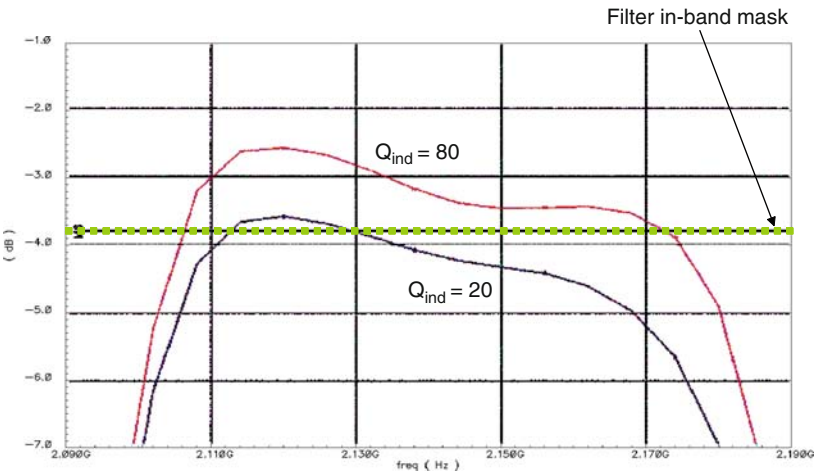


Fig. 12.13 In-band BAW filter response variation with respect to $Q_{inductor}$

System studies have shown that the quality factor of the parallel inductor is directly impacting the in-band losses of the filter, which is a critical specification. Figure 12.13 is depicting this phenomenon and is providing specifications for this inductor. The horizontal marker is the lower limit of the in-band filter mask, thus the quality factor of these parallel inductors should be superior to 80.

The filter is physically implemented by using the SiP co-integration between a Silicon 0.25 μm SiGe BiCMOS process and a stand-alone SMR BAW process, the two of them being interconnected using the flip-chip bumping method.

12.5.1 BAW Tuning Cell Implementation

Table 12.1 concentrates the major trade-offs for the Silicon integration of a high-Q inductor-like circuitry for frequencies around 2 GHz. A Si integrated spiral in the BEOL classically meets no more than 20 as a Q-factor. An active inductor implemented using the gyrator technique [7] shows high power consumption and noise and linearity issues for the given application. The final choice went towards a spiral inductor with Q-enhancement circuitry.

Figure 12.14 presents the proposed circuit for a Q-enhanced inductor. Part of the losses of the spiral BEOL inductor are compensated by a current controlled negative resistor implemented using the gyrator technique. A 4 nH inductor is required for resonators that present a 1.37 pF C_o for a 2.14 GHz filter center frequency (i.e. a 50 Ω characteristic impedance). The tuning cells on the parallel and the series arms use identical inductors.

Table 12.1 Trade-offs for the integration of a high-Q inductor (to be used in the schematic from Fig. 12.12)

	Integrated spiral inductors	Active inductance [7]	Q-enhanced inductors
Pros	Easy to implement no power consumption	Low area high Q (tunable)	High Q (tunable)
Cons	Large area low Q-factor	Prohibitive power consumption noise and linearity performances	Large area power consumption

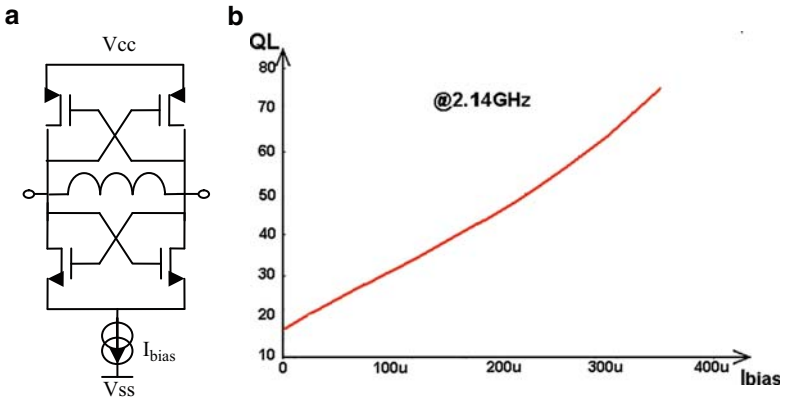


Fig. 12.14 (**a**) Q-enhanced inductor. (**b**) Variation of Q vs. I_{bias}

This Q-enhancement circuitry allows reaching a Q-factor of 80 for a current consumption of 350 μA. The equivalent negative resistance compensates the resistive part of the inductor in a narrow frequency band. An NMOS cross-coupled pair provides resistive impedance. Thus, adding a second PMOS cross-coupled pair allows increasing the negative resistance magnitude for the same tail current [8]. This biasing current controls the Q-factor of the inductance-like equivalent circuit (see Fig. 12.14b). In order to prevent the oscillations of the shunt Q-enhanced inductor and the BAW resonator, the negative resistance magnitude must be less than the resistive part of the impedance at its characteristic frequencies.

For a better distribution of the parasitic capacitances, the circuitry realizing a variable capacitance has been placed after the couple (inductor, BAW resonator), as depicted in Fig. 12.15. In fact, the inductance parasitic capacitances (around 300 fF) create a capacitive divider if the varactor is placed before this couple. In addition, the PMOS cross coupled pair fixes the common mode that we exploit to bias the varactor. This permits to eliminate one of the coupling capacitances needed in series with the varactor. Thus, the varactor block consists of a Metal Insulator Metal capacitor of 25 pF, a 100 kΩ poly resistor and an NMOS varactor. Figure 12.15a shows the final tuning cell implementation while Fig. 12.15b exhibits its simulated impedance for different varactor tuning voltage.

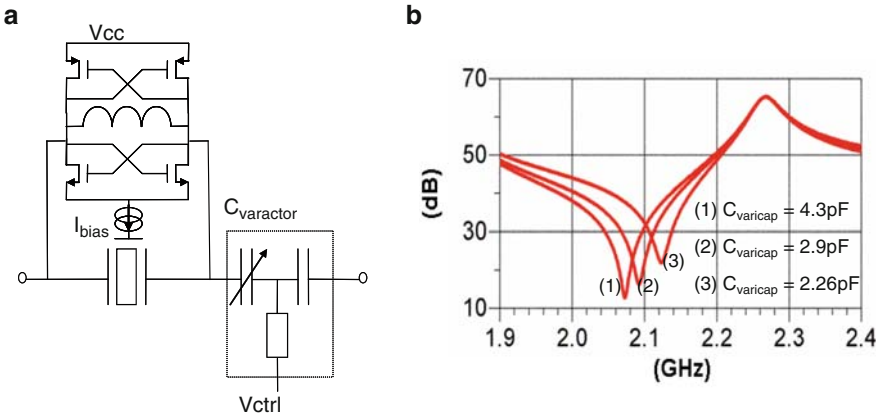


Fig. 12.15 Final tuning cell implementation: (**a**) scheme and (**b**) its impedance for different values of V_{ctrl}

Each BAW resonator of the synthesized BAW filter is replaced by its respective tuning cell. Now, the filter tunability becomes function of the tuning range of the varactor while the transfer function depends on only one control voltage (V_{ctrl}).

12.5.2 BAW Filter Implementation

As expected, 50 Ω scattering parameters analysis exhibit good results on this filter. Indeed, in-band insertion losses are reduced by 1 dB thanks to the Q-enhanced inductance. The tuning cell permits to correct 1.4% of shift on the piezoelectric layer as shown in the Fig. 12.16.

To demonstrate the validity of this tunable filter, two chips have been implemented: the first one contains the 8 SMR BAW resonators processed at CEA-LETI whereas the second one is a 0.25 µm SiGe BiCMOS 0.25 µm SiGe process from STMicroelectronics. They are assembled using a flip-chip bumping process. However, the flip-chip assembly brings some limitations on the insertion losses. Indeed, bump pads do not permit to optimize the filter size and add extra interconnection losses. Their design rules impose spatial constraints in order to balance the pressure during the assembly.

Besides, this filter architecture is composed of 8 high-value on-Silicon inductances whose size is rather large. In order to avoid the self coupling, a minimum spacing is effectively required hence contributing to the increase in the filter size. Figure 12.17 is giving a cross-section of the system physical co-integration, while Fig. 12.18 is showing a picture of the assembled circuit and a layout of the SMR BAW die which is flipped on the top of the Si die.

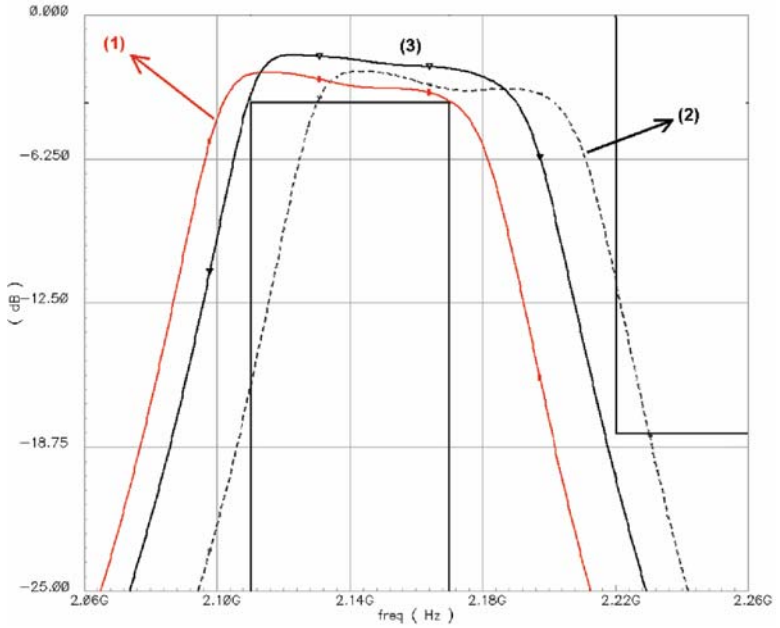


Fig. 12.16 Simulated BAW filter response: (1) Nominal filter response, (2) filter response with BAW resonator presenting 1.4% shift of the piezoelectric layer thickness, and (3) filter response with BAW resonator presenting 1.4% shift of the piezoelectric layer thickness after correction by the varactor control voltage

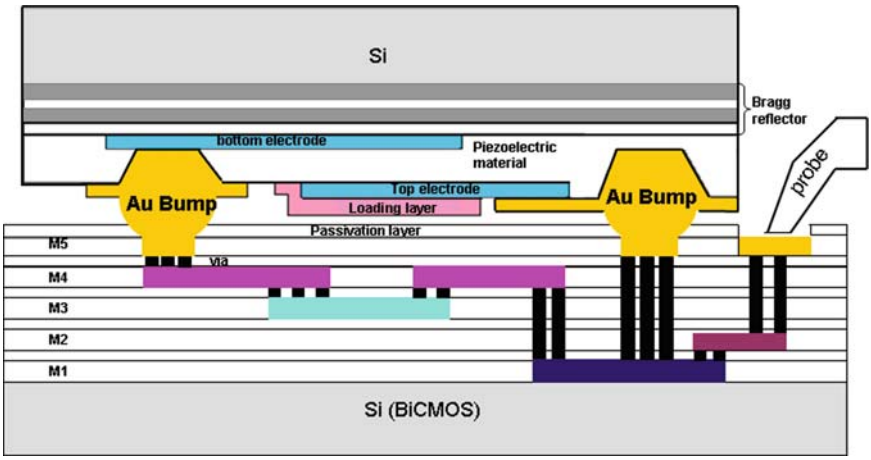


Fig. 12.17 Technology stack for the circuit co-integration

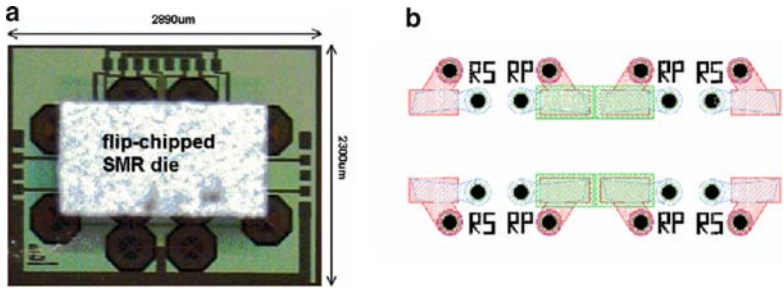


Fig. 12.18 (**a**) The SiP co-integration BAW filter photo-micrograph, (**b**) layout of the SMR BAW die

Table 12.2 SMR BAW electrical characterization

SMR resonators	C_o (pF)	Impedance @ 2.14 GHz (Ω)	Parallel SMR f_s (GHz)	Series SMR f_s (GHz)
Targeted	1.37	54	2.0528	2.1158
Measured	1.77	42	2.050	2.124

12.5.3 BAW Filter Measurement Results

The electrical characteristics of the SMR BAW resonators used in the filter are summarized in the following table:

The in-band ripple is dependent on the ratio of filter's branches impedances and also on the phase difference, as stated in the first part of this chapter. The measured filter in-band ripple (1.5 dB) is larger than the simulated one (0.8 dB) and may be explained by the slight discrepancies in Table 12.2. The measured Maximum Available Gain is reported on Fig. 12.19. An out-of-band rejection of 28 dB has been measured over a wide frequency band. The first prototype exhibits an extra 2 dB insertion loss compared to the post layout simulations.

Several factors may be responsible for this in-band gain degradation. First of all, the coupling between BiCMOS on-chip inductors is enhanced by the metallic plate of the flip-chipped die. Secondly, the large chip area implies important parasitics on the access lines. Finally, the bump access resistance degrades the BAW resonator's Q-factor.

The filter's center frequency may be tuned over a 0.3% relative frequency band that corresponds to a correction of a 0.6% error on the piezoelectric layer. Moreover, two notch frequencies due to mismatch between impedances of series and parallel branches appear near the pass-band.

The other electrical measured performances of the presented filter are given in Table 12.3.

Further optimizations should considerably improve performances of such filters in the future. Extra efforts on the layout of a second tunable BAW filter has lead to constant and lower insertion losses in the filter's bandwidth. Indeed, an improved

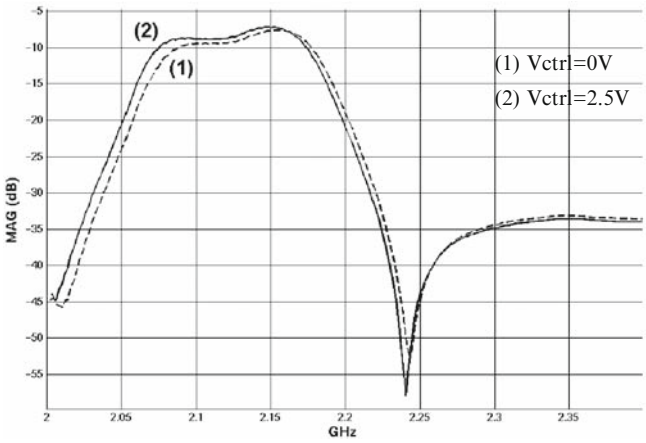


Fig. 12.19 Measured Maximum Available Gain (MAG) of the BAW tunable filter

Table 12.3 BAW filter measured performances

Excess noise factor (dB)	IIP3 (dBm)	f_o tunability	Power consumption	Si die area (mm^2)
0.2	35	0.3%	2.8 mA × 2.5 V	6.65

filter structure using less inductors, an increase in the BAW resonator's Q-factor and above-IC integration of such filters will also minimize losses due to the parasitic elements.

12.6 Tuning Circuitry for BAW Filters

12.6.1 Preliminary Discussion

In the classical literature about integrated filters tuning, two methods are generally presented: the direct and the indirect one. The direct tuning method (cf. Fig. 12.20a) implies that the electrical bloc to be tuned is taken off the signal path during the tuning period [9–11]. This is incompatible with mobile communication standards with time division multiplexing mode. Thus, such a calibration becomes inherently impossible to conceive for numerous standards unless the calibration can be operated before starting all communication phase.

The master/slave technique corresponds to an indirect tuning method (cf. Fig. 12.20b). This one has particularly been used with Gm-C filters [12]. One of the approaches used for the master/slave technique is to lock a given device referred to as the master circuit with respect to a fixed time reference using the well-known frequency synthesis methods like PLLs and to use the generated quantity

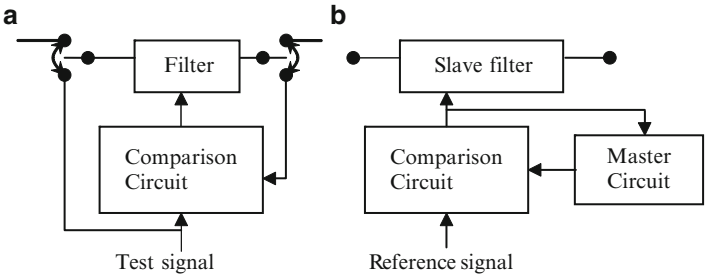


Fig. 12.20 (**a**) Direct tuning principle. (**b**) Indirect tuning principle

(usually the control voltage of the VCO) to tune the slave circuit, which has to be composed of the same basic elements as the master circuit. Parasitic elements added by the tuning circuitry can make the master cell environment vary with respect to the slave cell environment and thus generate a shift in the master cell impedance, thus providing a bad correction. The master circuit has to be matched to the slave circuit directly (i.e. exactly the same elements) or homothetically (i.e. there is a constant factor between the values in the Master and the Slave [11]).

Several tuning strategies may be foreseen for the automatic tuning of the *tunable* BAW filters. The choice of the master and the slave circuit turns out to be crucial. Obviously, for reason of size, the filter cannot be duplicated. In this case, the different decision criteria can be the following ones:

- A pass-band filter has its central frequency defined when its phase is zero. A direct tuning method could be exploited to detect the phase difference between the output signal and the test signal feeding the filter's input.
- As a control voltage allows shifting the BAW resonator impedances with the same amount, one of the resonant frequencies of one of the tuning cells can be also tuned and controlled by an indirect tuning system.
- Finally, as the low pass-band cut-off frequency of the proposed tunable filter is defined at the frequency at which series and parallel impedances are equal in magnitude but opposite in phase, a third solution consists of detecting impedance magnitude by an indirect tuning method.

In fact, as the insertion losses of the filter imply a shift between the frequency at which the filter phase is null and its central frequency, the first proposed solution turns out difficult to be implemented. Moreover, an extra complex loop would be needed to correct such phase shift. On the other hand, distortion harmonics in the reference signal will also create phase error by interfering with the phase detector [10]. Therefore, only the two others solutions will be discussed in the following part. They are applied to the BAW tunable filter given in Fig. 12.12.

12.6.2 Indirect Tuning Method I: PLL with a VCO as Master Cell

Among the indirect tuning methods, the tuning of the resonant frequency of one of the BAW based resonators (in series or in parallel branches) can be implemented. The BAW resonators can be exploited in a VCO structure and inserted in a PLL. The type of the VCOs (series or parallel resonant tank) will define the resonant frequency to use. However, according to the operating principle of the studied filter, it turns out to be mandatory to use the series resonant frequency (see Fig. 12.21a).

The efficiency of this tuning technique implies several operation mode constraints. First of all, the oscillation amplitude has to be controlled in order to place the master and the slave circuit in the same operating mode. The negative resistance of the Q-enhancement inductors can generate non-linearity and disturb the locking of the PLL. Moreover, the VCO has to be representative to the filter sensitivity to process and temperature frequency deviations. The tuning cell exhibits two resonant frequencies (see also Fig. 12.10). The parasitic one is determined by the inductor value and by its two neighbor capacitors (BAW resonator C_0 and the varactor). This

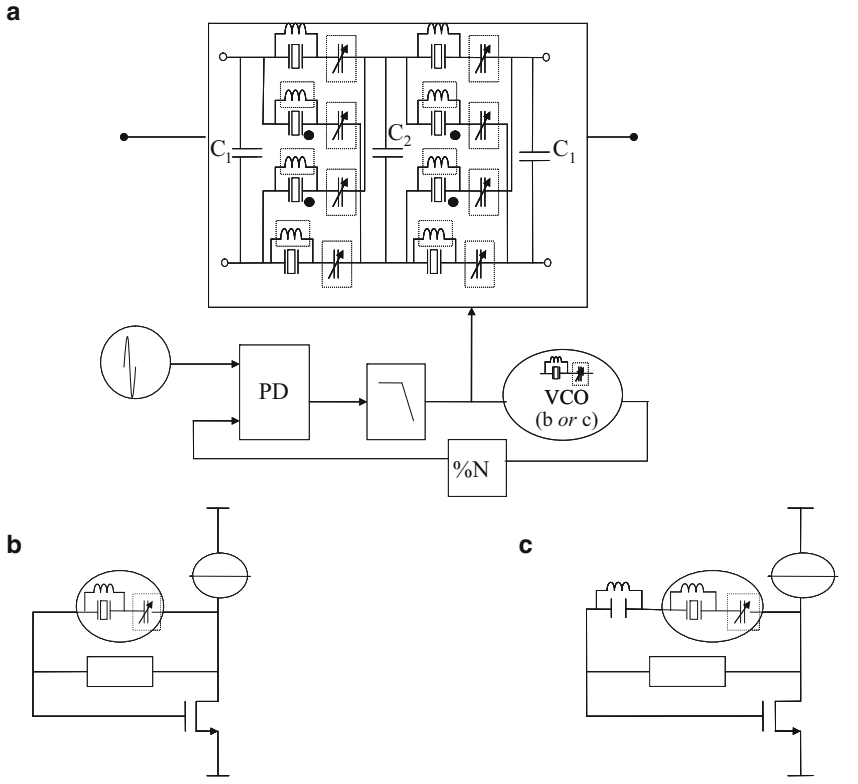


Fig. 12.21 Generic example of a master/slave tuning circuit (**a**) with a matched Pierce VCO on the parasitic resonant frequency (**b**) and on the useful series resonant frequency (**c**). [See also Fig. 12.10]

lower resonant frequency is very attractive for reaching lower power consumption. Moreover, the plate electrode capacitor C_0 is dependant on the thickness of BAW resonators and thus could well characterize the main BAW process dispersions. The main drawback is that the resonant frequency of BAW resonators is also defined by all the other stacked materials. Therefore, it seems to be unavoidable to exploit the useful resonant frequency.

The series resonant frequency can be easily used by placing the tuning cell in the direct feedback of the VCO as can be done in a Pierce configuration VCO (see Fig. 12.21b). Thus, naturally, the VCO will oscillate at the frequency requiring the less energy (which is the parasite one) and an extra trap resonant circuit has to be used to force oscillation at the second desired resonant frequency (see Fig. 12.21c). However, this extra circuit inserts parasitic capacitances contributing to supplementary mismatches and thus to a quasi systematic tuning error.

This master/slave technique cannot be applied with this tuning cell but can be convenient with a tuning cell using a negative capacitance as the one presented in [5]. However, even in this case, oscillating at 2 GHz needs relatively high area transistors loading directly the tuning cell and shifting the oscillation frequency towards the effective series resonant frequency of filter's BAW resonators. Matching the slave filter to the master VCO at gigahertz frequencies is difficult. Their respective structures are very different and do not naturally match. Extra capacitors in the oscillator core give rise to frequency pulling which results in a non negligible tuning error.

An alternative solution is possible with an envelope detection to determine the characteristic frequencies of the filter.

12.6.3 Indirect Tuning Method II: FLL with Envelope Detection

For this tuning method, we plan to detect the frequency at which the two resonant structures from the series and parallel branches of the filter in Fig. 12.12 are equal in terms of impedance's magnitude. If we limit the detection frequency range, we are located in the frequency interval where both impedances have opposite phase. In this way, by using two gain blocks proportional to Z_s (impedance of the series tuning cell) and respectively to Z_p (impedance of the parallel tuning cell) and by injecting a signal at a reference/input frequency, we are able to compare the level of both impedances at this frequency (see Fig. 12.22) [13]. Associated to a low-pass filter, an envelope detector will provide this information. A true bit ($= 1$) will be generated by a comparator if the difference between Z_s and Z_p (noted Δ) is positive. Otherwise, a false bit ($= 0$) will be generated.

The comparator output controls a successive approximation register (SAR) associated to a DAC, which will adjust the tuning voltage to the appropriated value, i.e. where both impedances Z_s and Z_p are equal in magnitude. Indeed, the low frequency clocked SAR will increment the central frequency's control voltage by addressing the adequate bits of the DAC block and using a dichotomy tuning

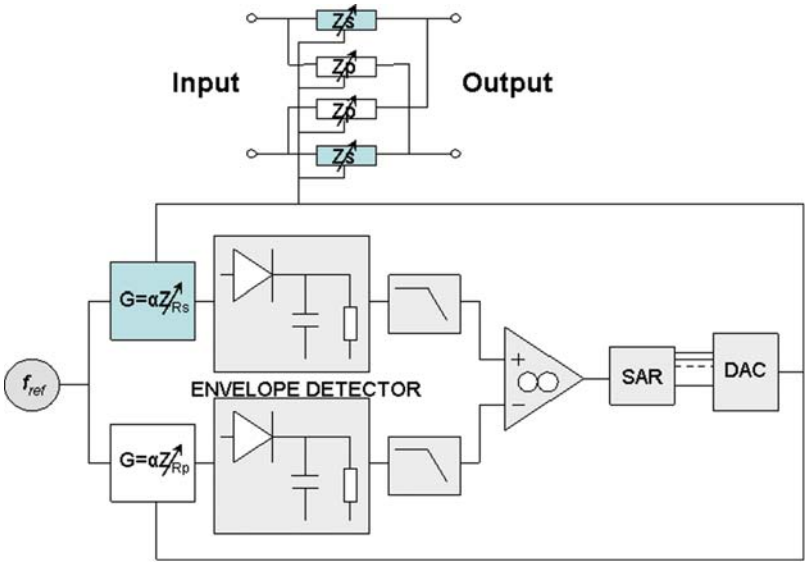


Fig. 12.22 FLL tuning circuitry principle

law. The resonant structures characteristic frequencies will be shifted by the same amount. This tuning method will be applied till the sign of Δ changes. The final control voltage value to be applied to the filter has been thus found. Then, by injecting this control voltage to the filter's control voltage, the filter's central frequency is maintained close to the reference frequency. Moreover, a supplementary bit (CLEAR) may be added in the SAR module to launch the tuning of the filter. According to the requested application, it is possible to tune it continuously and automatically or only punctually.

Furthermore, the accuracy of the tuning circuitry is highly linked to the matching of the gain blocks proportional to Z_s and Z_p and the resonant structures used effectively in the filter. It also depends on the reference clock. Indeed, because of the time constant, the tuning circuitry needs to have time enough to reach the steady state and then to come to the good detection. The slower the clock, the more exact the decision. Finally, the DAC will define the steps of the control voltage and has to be adapted according to the required application.

12.7 Design of a Digital Tuning Circuitry for a BAW Tunable Filter

In this section, the practical implementation of the tuning circuitry given in the previous Sub-section 12.6.3 is presented [14]. As in the first design implantation, the circuit is obtained by the SiP flip-chip co-integration between a 0.25 µm SiGe BiCMOS die and a SMR BAW die (see Fig. 12.17).

12.7.1 Circuit Implementation

The designed tuning circuit is presented in Fig. 12.23a. The gain of the input structures is proportional to the impedances of each branch as given in the following equation:

$$G_{s,p} = \frac{Z_{s,p}}{R + Z_{s,p}} \tag{12.8}$$

where R is placed in the direct path and $Z_{s,p}$ are the grounded impedances.

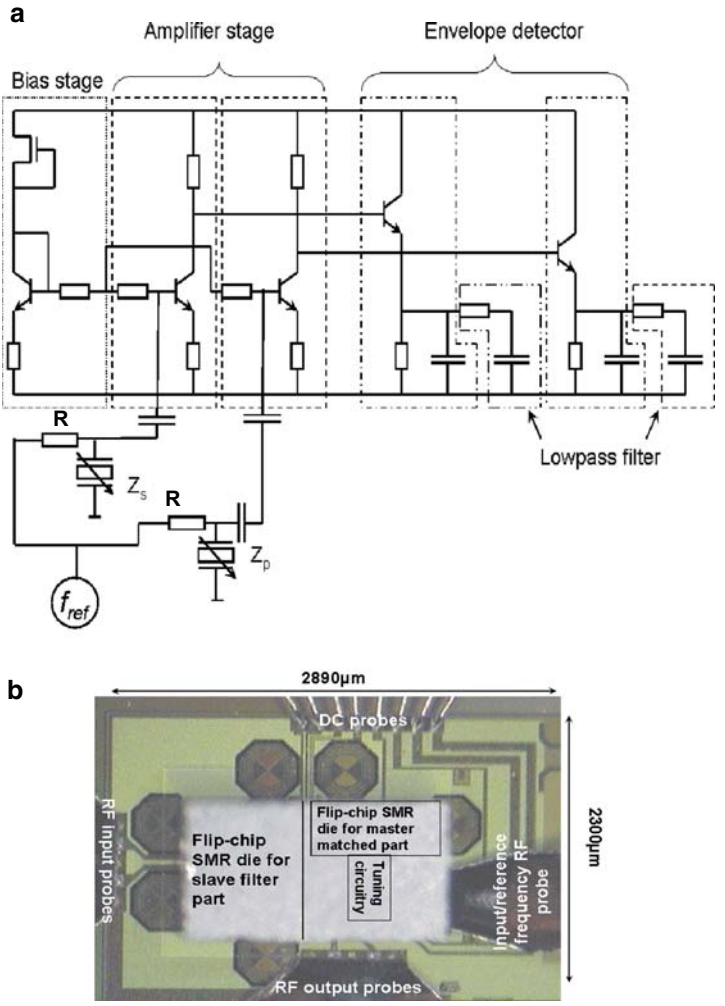


Fig. 12.23 (**a**) Down-converted part of the tuning circuitry and (**b**) microphotograph of the master/slave system

To match the slave to the master cell, Z_p and Z_s have to be loaded by the same capacitors seen by the respective slave impedances employed in the filter to be tuned. On the other hand, in order to increase the accuracy of the detection, the resulting output signal will be amplified before sampling its envelope magnitude.

The Fig. 12.23a exposes the first part of the tuning circuitry with its amplifier stage associated to the envelope detector. It allows reducing the constraints of high frequency design by down-converting the tuning operating frequency in the MHz frequency range. This pseudo-differential part has to be well-matched. Indeed, the error committed on the envelope detection is at the first order the same on the two paths reducing the error on the decision. This circuit part consumes 2.54 mA current under 2.5 V and exhibits a 3 dB gain. The envelope detector is directly dc coupled with the amplifier output. It exploits the PN junction of a common collector bipolar transistor as a diode (see Fig. 12.23a).

Then, an O.T.A. consuming 220 µA under 2.5 V has been designed as a comparator and provides the decision bit equal to 1 when the Z_s magnitude is greater than Z_p one and otherwise a bit equal to 0. This bit called RESULT will control the SAR whose time reference will clock the filter tuning. For the proposed SAR design, a single D-type flip-flop is used in each bit cell which functions both as sequencer and code register. This type of design is often referred to as the sequencer/code register design [15]. It consumes 40 µA and will be clocked from 1 MHz to 5 MHz. Furthermore, to reduce the locking time, at the beginning of the tuning sequence, the SAR is initialized to the mid-value of the control voltage interval, i.e. MSB = 1, all other bits are 0. The bits are then adjusted according to the comparator output value starting from the MSB going towards the LSB, the final value of the tuning voltage being within 1 LSB from the ideal value.

The slave filter is a one section lattice filter and has a structure similar to the one presented in Fig. 12.12. The BAW chip contains 6 SMRs (four resonators for the filter and two resonators for the master cell) in an AlN piezoelectric layer process provided by CEA/LETI whereas the second one is processed in a 0.25 µm SiGe:C BiCMOS technology from STMicroelectronics. The total master/slave chip area is 6.5 mm^2 whereas the tuning circuitry Si footprint is less than 0.15 mm^2 and a microphotograph of the SiP assembly is given in Fig. 12.23b.

12.7.2 Measurement Results

For testability reasons, the reference frequency has not been implemented on chip and thus is provided by an external source. This source has been swept from 2.07 to 2.09 GHz with a step of 100 kHz and a 1V p-p amplitude. As depicted in Fig. 12.24a, the displayed results represent the behavior of the Slave BAW filter control voltage versus the external reference frequency whereas the variation of the corresponding central frequency (f_0) is also drawn. The measured BAW filter presents a constant 104 MHz bandwidth and its central frequency is controlled on a 10.5 MHz frequency range. Finally, the obtained error is about 3.6 MHz which is less than 0.2% of the

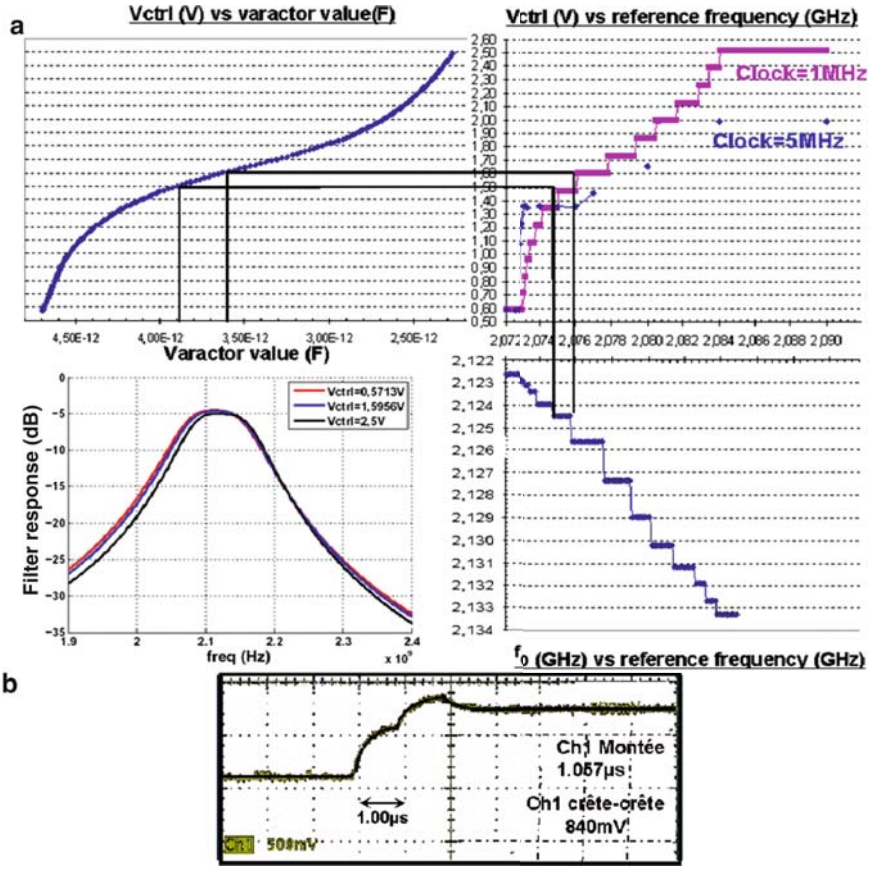


Fig. 12.24 Measurement results for the tuning circuitry: (**a**) Behavior of the BAW filter (control voltage) with respect to the external f_{ref}; (**b**) V_{ctrl} settling

filter central frequency whereas the tuning step is 100 kHz. Furthermore, the tunability is facing the non linear capacitance variation of varactors toward its control voltage. Implementing a non linear step for the DAC could sort out such an issue. Furthermore, the control frequency range is in this implementation limited by the tuning varactor range but can be made larger by the use of a matrix of switchable capacitors.

Two reference clock frequencies (1 and 5 MHz) have been tested. A quasi systematic error is committed by the use of the faster clock because the decision has been done before achieving the steady state. The Fig. 12.24b capture has been done for a 2.083 GHz input frequency and the measured signal is the BAW filter control voltage.

One can observe the dichotomy function of the SAR-DAC stage. The rise time is defined by the product of the decoupling capacitor's varactor to load and its decoupling resistors. It actually determines the clock frequency to use. Nevertheless,

Table 12.4 Tuning circuitry electrical performances

Parameter	Value
Tuning precision	100 kHz
Error on the BAW filter central frequency	<0.2%
Settling time for a 4-b tuning circuit clocked at 1 MHz	4 μs
Power consumption	(2.54 . . . 4 mA) × 2.5 V
Tuning circuit area	0.15 mm^2

the settling time of the tuning circuitry is function of the error between the targeted central frequency and the effective one. This value is included between 2 and 4 μs. Finally, depending on the applied correction voltage, the tuning circuitry consumes from 2.54 to 4 mA from a 2.5 V power supply. However, if a punctual correction is sufficient, the tuning circuitry can be turned off after storing the control voltage.

Table 12.4 is resuming the electrical performances of the presented tuning circuitry.

12.8 Conclusions and Perspectives

This paper has presented a novel category of analog/RF integrated filters, which take benefit of BAW piezoelectric resonators. BAW resonators may be considered as parented to a kind of mid-Q quartz resonators (their Q-factor is only around 1,000), but they show the very interesting feature that their processing permits to build on one chip filters or any other structures of designer-defined ad-hoc layout. Moreover, their size (and chip height) is compatible with the ones found usually in analog ICs. Thus, these BAW chips may be co-integrated with regular Silicon dies by using for example flip-chip bumping assembly. Once these passive high-Q BAW resonators are enhanced with active Silicon "intelligence", we can tune and correct with a good precision process and temperature variations for these devices.

This paper has first presented a 2 GHz *tunable* BAW filter, and details have been given on a novel structure of a BAW *tunable* cell by electrical means. In this way, we have validated the feasibility of synthesizing and designing tunable BAW band-pass filters able to compensate the BAW piezoelectric process and temperature dispersions. Measurements have generally shown good agreement with post-layout simulations. Another kind of BAW tuning cell using negative capacitor-like circuitry [5] may permit to lower the Si die area by avoiding the use of any Si integrated inductors, and this permits also to compact the layout and thus reduce the interconnections parasitic effects.

Moreover, this work offers other new perspectives on filters design. For example, let's consider the BAW *tunable* cell depicted in Fig. 12.10 and a filter implemented as in Fig. 12.12 and let's impose different values for the inductors or varactors from the series resonators and the parallel ones. In this case the initially designed 2 GHz

band-pass filter shows also an extra resonant frequency which can be exploited in order to obtain a second pass-band filter at lower frequencies [16].

In order to automatically correct frequency deviations, we need to associate a tuning circuitry to these BAW tunable filters. Therefore, a master/slave technique which allows to digitally tune a tunable BAW filter's central frequency with respect to a reference clock has been proposed also in this paper. This system is aimed for correcting process and temperature dispersions and also for tuning all type of lattice filters accurately. Instead of relying on the sometimes complex PLL tunability, the whole flexibility has been moved to the proposed master/slave scheme, which consists of envelope detection. The control frequency range is limited by the tuning varactor range and can be made wider by the use of a matrix of switchable capacitors. This kind of schematic may be suitable also for reconfigurable applications if a convenient filter is designed.

Regarding the physical implementation, the intimate SiP BAW/IC co-integration has been demonstrated with a 0.25 μm 2.5 V BiCMOS process and BAW SMR resonators. From now on, individual BAW devices may be seen as intrinsic parts of an IC process like inductors or capacitors. Thus, we may think of redefined circuit architectures taking benefit of the large Q factor of such resonators. Typical examples of system blocs where BAW devices are attractive are namely oscillators [17–20] or again any other kind of filters.

References

1. Anatol I. Zverev, "Handbook of Filter Synthesis", Wiley, 1967.
2. R. G. Kinsman, "Crystal Filters: Design, Manufacturing and Application", Wiley, A Wiley-Interscience Publication, April 1987, p 52–56.
3. C. Tilhac et al., "A Bandpass BAW-Filter Architecture with Reduced Sensitivity to Process Variations", IEE J Analog VLSI Workshop, 2005.
4. S. Razafimandimby, A. Cathelin, J-F Carpentier, D. Belot, "Electronic Circuit Comprising an Adjustable Resonator", US granted patent US7187240, Mar. 2007.
5. C. Tilhac et al., "A Tunable Bandpass BAW-Filter Architecture Using Negative Capacitance Circuitry", in Proc. IEEE Radio Frequency Integrated Circuits Symposium 2008, Atlanta, June 2008.
6. S. Razafimandimby et al., "An Electronically Tunable Bandpass BAW-Filter for a Zero-IF WCDMA Receiver", Proc. ESSCIRC 2006, Montreux, Sept. 2006.
7. B. Nauta, "A CMOS transocnductance-C Filter Technique for Very High Frequencies", IEEE J. Solid-State Circuits, vol. 27, no. 2, Feb. 1992.
8. A. Cathelin, S. Razafimandimby, J-F Carpentier, D. Belot, "Electronic Circuit Comprising a Resonator to be Integrated Into a Semiconducteur Product", US granted patent US20050174198, Sept. 2008.
9. K.A. Kozma, D.A. Johns and A.S Sedra, "An Approach for Tuning High-Q Continuous-time Bandpass Filters", IEEE International Symposium on Circuits and Systems, vol. 2, pp 1037–1040, May 1995.
10. F. Krummenacher, N. Joehl, "A 4 MHz Continuous Time Filter with On-Chip Automatic Tuning", IEEE J Solid-State of Circuits, vol. 23, no 3, pp 750–758, June 1988.
11. A. Cathelin, "Tuning and Reconfiguration for Analog Filters", Tutorial at Filter Design Tutorial Course, ESSCIRC – ESSDERC 2008, Edinburgh, Sept. 15, 2008.

12. Y. P. Tsividis, “Integrated Continuous-Time Filter Design An Overview”, IEEE J Solid-State Circuits, vol. 29, no. 3, pp. 166–176, Mar. 1994.
13. S. Razafimandimby, A. Cathelin, A. Kaiser, “Tuning Circuitry for Lattice Filter”, FR0706161, FR filing patent application, Sept. 2007, US12203003, US filing patent application, Sept. 2008.
14. S. Razafimandimby et al., “Digital Tuning of an Analog Tunable Bandpass BAW-Filter at GHz Frequency”, in Proc. ESSCIRC 2007, Munich, Sept. 2007.
15. T. Anderson, “Optimum Control Logic for Successive Approximation A/D Converters”, Comput. Des., pp. 81–86, July 1972.
16. S. Razafimandimby, A. Cathelin, A. Kaiser, “The Multiple Band RF BAW Filter”, US patent publication US20080024244, Jan. 2008.
17. S. Razafimandimby et al., “A 2-GHz 0.25-μm SiGe BiCMOS Oscillator with Flip-Chip Mounted BAW Resonator”, in Proc. ISSCC 2007, San Francisco, Feb. 2007.
18. P. Vincent et al., “A 1 V 220 MHz-Tuning-Range 2.2 GHz VCO Using a BAW Resonator”, in Proc. ISSCC 2008, San Francisco, Feb. 2008.
19. M. Aissi et al., “A 5.4 GHz 0.35 μm BiCMOS FBAR Resonator Oscillator in Above-IC Technology”, in Proc. ISSCC, San Francisco, Feb. 2006.
20. B.P. Otis et al., “A 300-μm 1.9 GHz CMOS Oscillator Utilizing Micromachined Resonators”, IEEE J Solid-State Circuits, vol. 38, no. 7, July 2003.

Part III
Multi-mode Transmitters

The third part of this book is on RF circuits towards multi-mode transmitters. In the never ending story of SoC, one of the missing blocks towards multi-mode transmitters is the power amplifier. Especially the power efficiency for both base station and handset applications becomes a challenge. For that in this part techniques towards efficient transmitters, if possible for multi-mode applications are discussed.

The first paper, of Earl McCune, starts from the point of efficiency for the realization of transmitters. Instead of using the traditional linear amplifiers with low efficiency, an introduction of polar modulation techniques is discussed. The different problems with some solutions to both the AM–PM and AM–AM conversion are introduced. It is shown via realizations that modern applications, from Bluetooth up to IEEE802.11 a/g can be realized with those techniques.

The second paper, of Bo Berglund, addresses radio base station power amplifiers. The power range requires GaN or LDMOS devices. Since also the losses in the remaining filters are becoming important, alternative techniques are explained. Polar modulation techniques are described where the high frequency content of the envelop is delivered by a linear device while the bulk of the power (and low frequency) is delivered by a switching stage (very often called the beauty and the beast). Interesting techniques such as semi-amplitude modulation in order to relax the remaining filters are introduced.

The third paper, of Jan van Sinderen, addresses integration of multi-mode transmitters in CMOS. Different applications require attention to unwanted signals out of the band of interest. For that voltage modulators in combination with 25% LO generators are used, showing improved performances towards noise, ACLR and LO leakage.

The fourth paper, of Patrick Reynaert, describes the challenges for mobile terminal CMOS power amplifiers. Different topologies, such as polar modulation, out-phasing and power combining techniques are compared and analyzed. The trends towards more digital signal processing in CMOS allows the exploration of digital polar techniques where sigma-delta techniques are combined with polar switching power amplifiers.

The fifth paper, of Antoine Frappé, discusses the signal generation. Also in those building blocks, to the need of flexibility in multi-mode transmitters, the trends is also towards more digital preprocessing. The major drawback of the resulting

sigma-delta techniques is the shift of the problems to the remaining blocking/duplex filters. Adaptive placement of complex poles and zeros are discussed as a solution in order to relax the filters.

The last paper, of Henrik Sjöland, discusses cartesian switched mode architectures, based on LINC, PWM and delta sigma RF pulse with modulation. Especially the digital cartesian delta-sigma processing before going to polar PWM modulation shows improved performances (or relaxed filter specs) out the frequency band. It allows a reduction over a 100 MHz band with less than −50 dB adjacent channel power.

Michiel Steyaert

Chapter 13
Multimode Transmitters: Easier with Strong Nonlinearity

Earl McCune

Abstract Traditionally, multimode transmitters are approached from a linear circuit approach, and then significant work follows to improve their abysmal initial energy efficiency. Here we consider the reverse procedure, where the starting point is the circuit with greatest energy efficiency – a switch – surrounded by system architectures which result in exactly the same output signals. This latter approach necessarily leads to polar signal processing and polar modulation techniques.

13.1 Introduction

The need for building radios which simultaneously perform with excellent linearity while providing very high energy efficiency has never been more important. The people of the world have become familiar with the battery life and low operating temperature of GSM handsets, and want these characteristics to be maintained in the coming era of HSPA and OFDM signals. Our problem as designers is that the conventional trade-off that efficiency can be improved by accepting signal distortion is no longer available. As the industry adopts signals with ever higher order, these signals are less tolerant of any distortion. This is clearly evident in the change in the EVM specification from 17.5% for HPSK used for UMTS down to near 3% for OFDM used for WiMAX.

Legacy is another issue driving mobile device design. New communication systems are being turned on much faster than old systems are being shut down. And there is not one system type that provides complete coverage for all desired services. This means that to appear to provide complete coverage for all services to any mobile device customer, the mobile must be capable of adapting to the many system types that exist around the world. We must provide multimode operation.

E. McCune (✉)
Panasonic (retired), 2383 Pruneridge Ave, Santa Clara, California 95050, USA
e-mail: emc2@alum.calberkeley.org

A.H.M. van Roermund et al. (eds.), *Analog Circuit Design: Smart Data Converters, Filters on Chip, Multimode Transmitters*, DOI 10.1007/978-90-481-3083-2_13,
© Springer Science+Business Media B.V. 2010

Still, efficient use of battery energy is vital. This work is motivated by an interest in reversing the standard design process in an attempt to get around the traditional trade-off between energy efficiency and circuit linearity: choose one. In this approach we choose to begin instead with the most energy efficient circuit available, which is a switch. This circuit actually suppresses all linearity. Starting with this circuit we develop design approaches which do provide our desired multimode output signals, acting as if the transmitter has excellent linearity – even though all of the RF circuitry is extremely nonlinear.

13.2 Architecture

To access this maximum efficiency condition it is necessary to drive the amplifier sufficiently hard that the differential gain falls to nearly zero [1]. This operates the amplifier at the points shown in Fig. 13.1. Clearly this operation is beyond the linear and linearizable operating conditions, which emphasizes that this approach is fundamentally different from convention.

When operated at the conditions of Fig. 13.1, an appropriate model for the RF transistor is shown in Fig. 13.2. This model also holds for the final stage of the power amplifier. In this case the transistor is no longer regulating the current flowing through the load. Instead, it only selects when current will flow, or not. The actual amount of load current is controlled by the environment around the amplifier stage.

All signals of interest to modern mobile communications have two signal dimensions. These dimensions are known as in-phase (I) and quadrature-phase (Q).

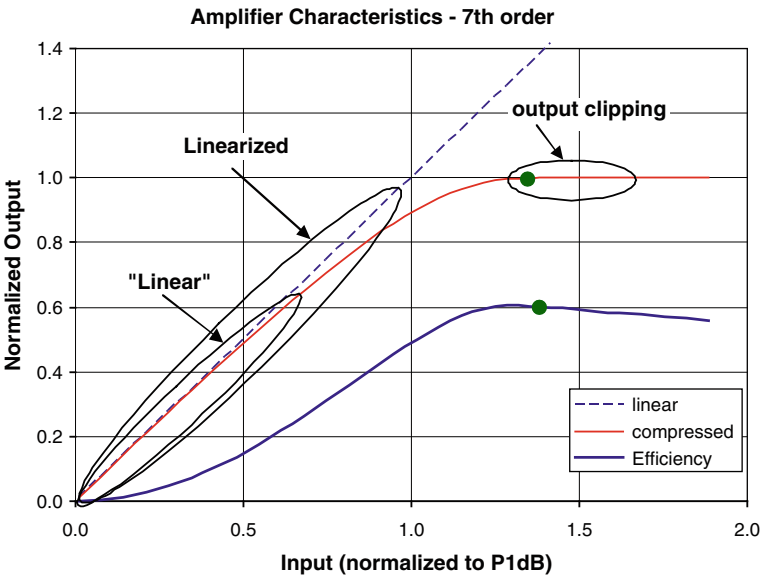


Fig. 13.1 Maximum efficiency condition of an amplifier

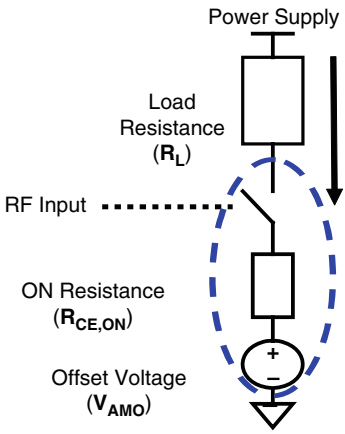


Fig. 13.2 Model of an RF power switch

However, without circuit linearity available these two dimensions are not useful in this circuit approach. This forces our design to shift the signal processing basis from I and Q to something appropriate for a switch.

The two independent parameters available to switch operation are (1) when the switch changes state, and (2) how much current flows through the switch when it is ON. Considering these in order, we first look at manipulating the time of switch action. Writing this time variation as $\Delta(t)$, we have

$$\cos(\omega(t + \Delta(t))) = \cos(\omega t + \omega \cdot \Delta(t)) \tag{13.1}$$

We see that this is equivalent to any desired phase modulation when

$$\phi(t) = \omega \cdot \Delta(t)$$

Output signal magnitude is directly proportional to current flowing in the load, which is identical to the current flowing through the switch in the model of Fig. 13.2. So with this circuit structure we do have direct control of phase and magnitude: polar coordinates. With this inverted design approach, our signal processing must shift to magnitude (ρ) and phase (θ). Use of polar techniques is gaining popularity. [2, 3]

Before discussing design details, it is useful to note that the desired multimode operation works fine with this approach. Figure 13.3 shows six different signals, including OFDM, all produced from one single design following this strategy.

13.3 Design Issues

Assuring that the RF circuits are actually operating as switches is very important. The natural assumption is to follow the definition of output saturation. However a better approach is to review Fig. 13.2, and to see that the design is looking for the

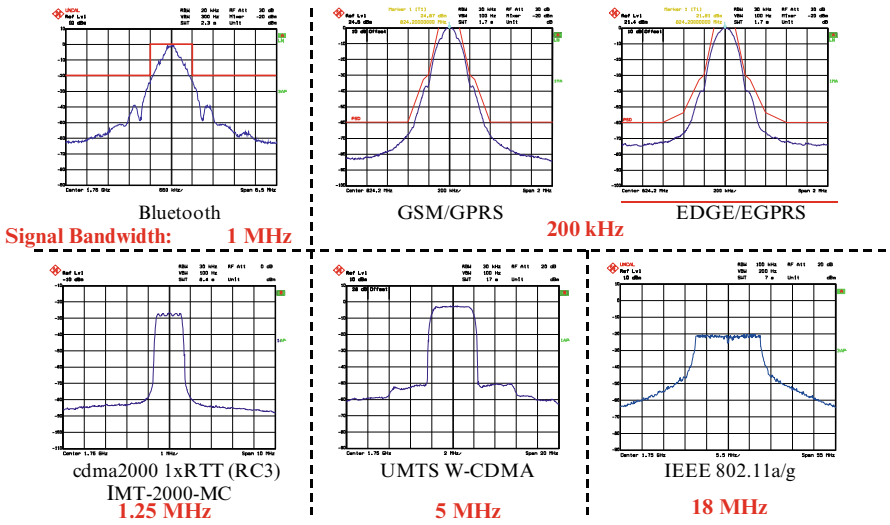


Fig. 13.3 Multimode operation achieved with this switch-based transmitter

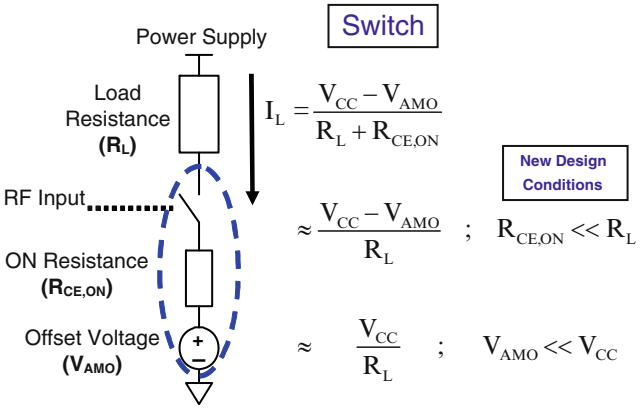
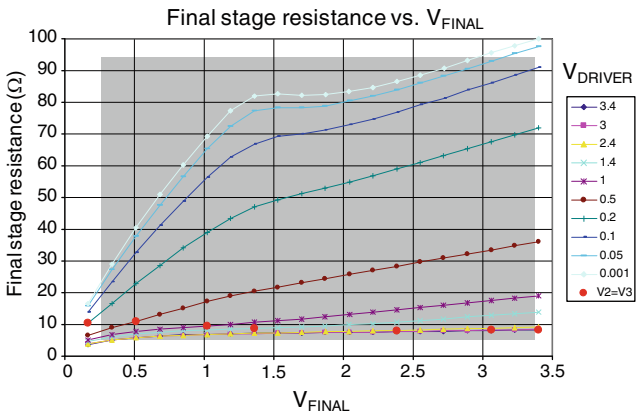


Fig. 13.4 Current flow control and design conditions for a switch-based circuit

transistor to not be regulating current through the load. This leads to the relationships of Fig. 13.4, including two new design conditions.

Following the relationships of Fig. 13.4, we introduce the concept of Stage Series Resistance (SSR). SSR is defined as the ratio of the current drawn from the power supply as the value of the power supply changes, evaluated at each value of power supply voltage. One result is shown in Fig. 13.5. By rewriting Ohm's Law in the form shown, we note that if the measurements of SSR plot as a line with non-zero slope, then the transistor is operating as a current source and not as a switch. The goal is to have the SSR value be constant across the variation range of the applied power supply.



$$R = \left(\frac{1}{I_{reg}} \right) V_{DD}$$

This device is acting as a CCS at all but the highest supply voltages

Fig. 13.5 Evaluating SSR for a switch-based circuit

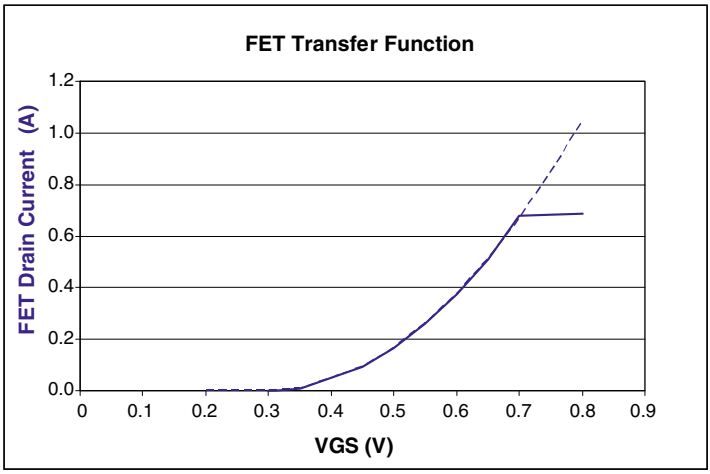


Fig. 13.6 Device current flow with current limiting

In order to emphasize a constant SSR behavior, one needs to pay close attention to the drive of each switch. Looking at Fig. 13.6, we see that when current through the switch device is limited by external means, there is a value of input drive beyond which there is no point to exceed. Further, as the amount of current required to flow through the switch is reduced, so too the amount of drive input to the switch is best reduced. Optimizing operation of these switching circuits includes varying the drive signal magnitude along with the desired output signal.

Control of the load current through the switch is moved outside the RF power stage and transferred to the envelope modulator (EM). There are many options available to realize the EM. By far the easiest and simplest is to use an agile linear voltage regulator. Called linear envelope modulation (LEM), this also results in the lowest

overall efficiency for this polar architecture. Even though the RF power device is operating at very high efficiency, this efficiency is not "visible" to the power supply because of the I*R losses in the linear regulator.

Efficiency can be dramatically improved by also including switching structures within the envelope modulator. Two prominent options are power-tracking switching (PTS) and envelope-tracking switching (ETS). Of these, ETS always provides the best overall efficiency. Surprisingly, for many important signals the efficiency gain of ETS over PTS is rather small.

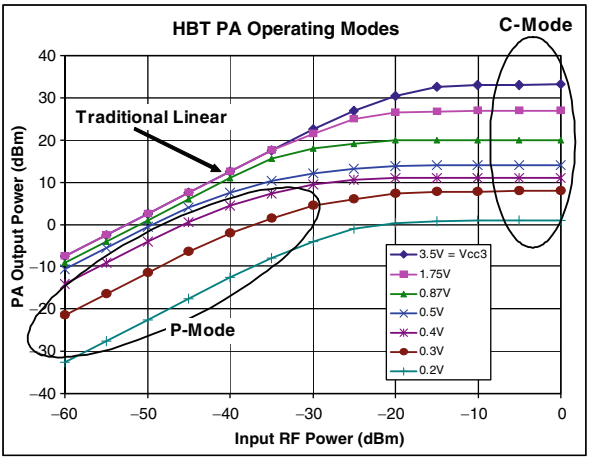
One very interesting side effect of switch-based operation is that there is no issue with circuit stability. A switch is unconditionally stable – it *cannot oscillate*. This is a direct result of the fact that a switched device exhibits no differential gain in either state. When OFF, clearly the device has no gain and cannot oscillate. When ON under the conditions defined above, there is essentially no differential gain available to support oscillation. It is extremely important to monitor input drive to assure this remains true. Should the input drive sag so that the transistor begins to regulate the load current, then differential gain increases dramatically and oscillations become likely.

Another interesting side effect of switch-based RF power generation has to do with broadband noise at the transmitter output. With any linear amplifier, there is a noise figure at the input that is subject to the gain of the amplifier. This yields the well known broadband output noise floor. But when the RF power device is operated as a switch, there is effectively no gain when power is generated. Correspondingly, the nature of the broadband output noise floor changes. Now the wideband output noise tracks the phase noise on the input signal. With a sufficiently low input noise floor from the phase modulator, the same performance is transferred to the final transmitter output.

Power control is a fundamental requirement of any mobile transmitter. This is especially important for any CDMA system. For this design approach, it is possible to achieve all the power control requirements for GSM, GPRS, and EDGE by direct scaling of the signal to the envelope modulator. To achieve the additional power control needed by CDMA systems, it is necessary to use an additional mode available from these switch-based circuits.

In Fig. 13.7 a set of curves illustrates the three major modes available from an amplifier design. These modes are summarized in the associated table. Conventional linear operation is characterized by the output signal being sensitive to the applied input signal power, and insensitive to changes to the power supply voltage. Switch operation, here also called compressed mode (C-mode), is the opposite: the output signal is sensitive to the applied power supply voltage, and insensitive to changes on the input signal power.

Figure 13.7 also shows a third available operating mode, here called product mode (P-mode). Here the output signal is responsive to both the input signal magnitude and the power supply voltage – but only when the power supply voltage is very small. This is not a linear operating mode, because there is proportional sensitivity to the applied power supply. The stage transistor is still not regulating the stage current, as it must for linear operation.



Operating mode can be defined by which input parameters the RF output is sensitive to

	Linear	C-Mode	P-Mode
$\frac{\partial Pout}{\partial Vcc}$	0	αVcc	κVcc
$\frac{\partial Pout}{\partial Pin}$	G	0	G

Fig. 13.7 Operating modes available from an amplifier

How this actually works is seen in Fig. 13.8. The RF transistor is operating as a resistor, in accordance with the model of Fig. 13.4, but here at the combination of low supply voltage *and* low input drive signal the device ON resistance varies proportional to the drive signal magnitude. In this case the input signal is still a fixed magnitude (constant envelope) and contains the desired phase modulation. The multiplication process which imposes the desired signal envelope by variation of the stage supply still operates as before in fully compressed mode.

When operating in Product Mode the range of power supply variation is small, but importantly it may now be fixed at a convenient but appropriate level. Additional power control to very low output levels is now achieved by attenuating the phase modulated constant envelope drive signal. This is the opposite situation from conventional switching operation, but it is still a polar operation and is definitely not linear. And as before, the control of signal envelope is still with the polar envelope modulator, and envelope modulation occurs at the same circuit location at all output powers.

13.4 Performance Measurements

The first measurements presented are of the case temperature rise observed at the power amplifier (PA) for GPRS and EGPRS operation when mounted on a normal handset circuit board. Duty cycle is varied from 1/8 to 8/8. In each burst the peak output power from the PA is +34.1 dBm, which is 2.5 W. Of course for GPRS this is also the average output power for each burst. When the EDGE 8-PSK modulation is used, the average power drops 3.2 dB to +30.9 dBm, greater than 1 W.

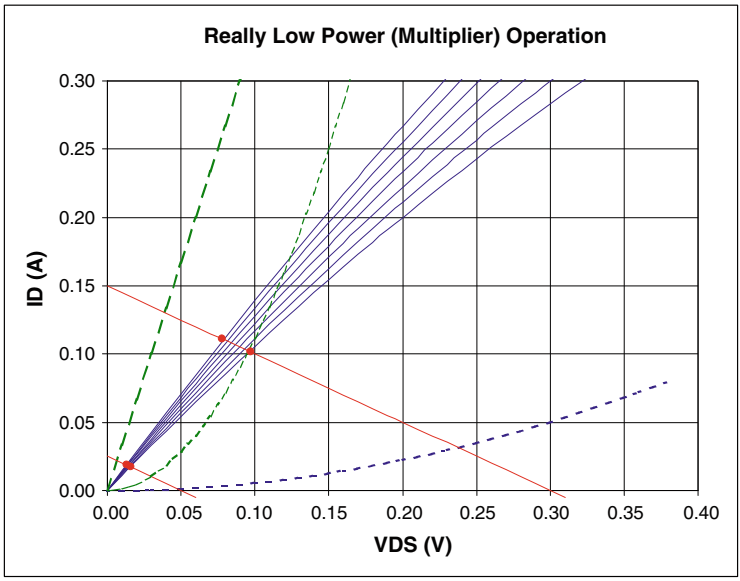


Fig. 13.8 Product mode operation is in the resistive part of the device characteristic curves. The device is not operating as a current source

With the same peak output power, the temperature rise for EDGE is lower than that for GPRS. At 50% duty cycle (4/8), the case temperature rise is 11°C for GPRS and +6 °C for EDGE at these power levels. This meets the original objective for low operating temperature. These results directly imply a very high energy efficiency.

Measurements of output signal quality and energy efficiencies are presented in Fig. 13.10. Both EDGE and WCDMA output signals are shown. Looking first at signal quality, the EDGE signal exhibits more than 10 dB of margin to the transmit mask specification at 400 and 600 kHz offset frequencies. This measurement is at the same power level used in the measurements of Fig. 13.9. This design uses the LEM architecture, so overall energy efficiency is not as high as possible. Still, the energy efficiency (PAE in this case) at full output power is 40%, much better than that available from a linear amplifier. The solid black line is the predicted PAE for LEM operation, and the black circles are measurements across several values of output power. The solid green curve corresponds to predicted efficiency of the RF device itself, and associated green circles are measurements of actual RF device efficiency.

For the UMTS signal, the signal quality at peak output power is excellent, showing −50 dB for ACLR5 and nearly −60dB for ACLR10. Corresponding efficiency predictions and measurements are very good. Further qualifications of this approach regarding manufacturing stability and yield are reported in [4].

One of the important issues noted for this strongly nonlinear circuit approach is to achieve very good inherent thermal stability from this transmitter. This stability with respect to wide temperature variations is a direct consequence of the design

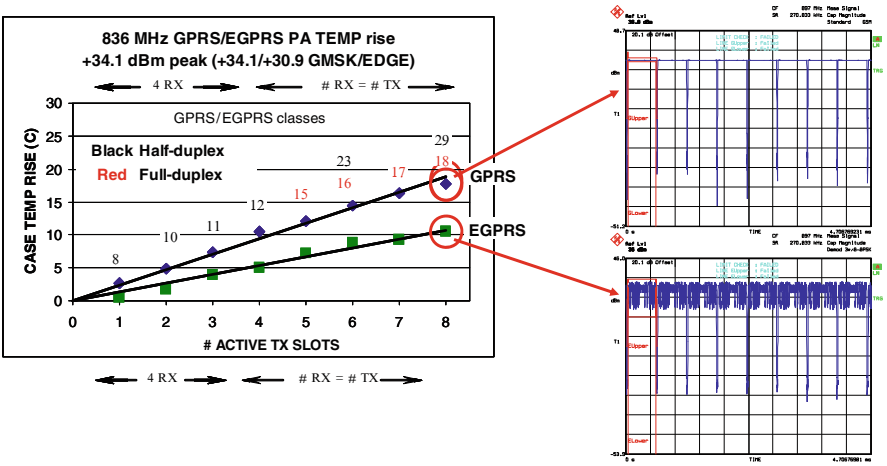


Fig. 13.9 Temperature rise measurements vs. transmitter duty cycle for GPRS and EGPRS at peak output power of 2.5 W

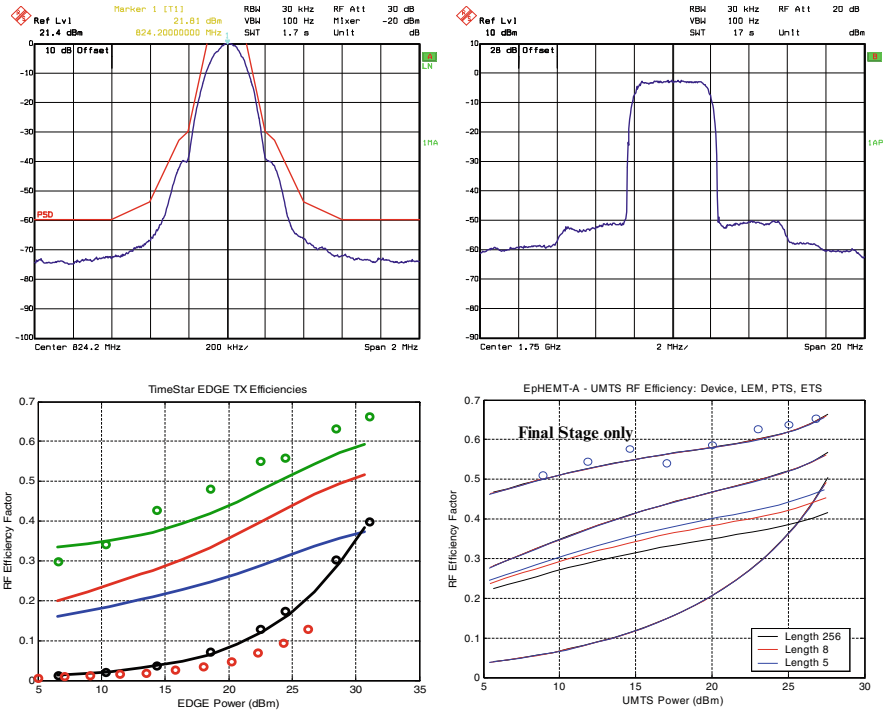


Fig. 13.10 Output spectral quality measurements for EDGE and WCDMA, along with corresponding efficiency measurements, across the top 20 dB of transmitter power control dynamic range

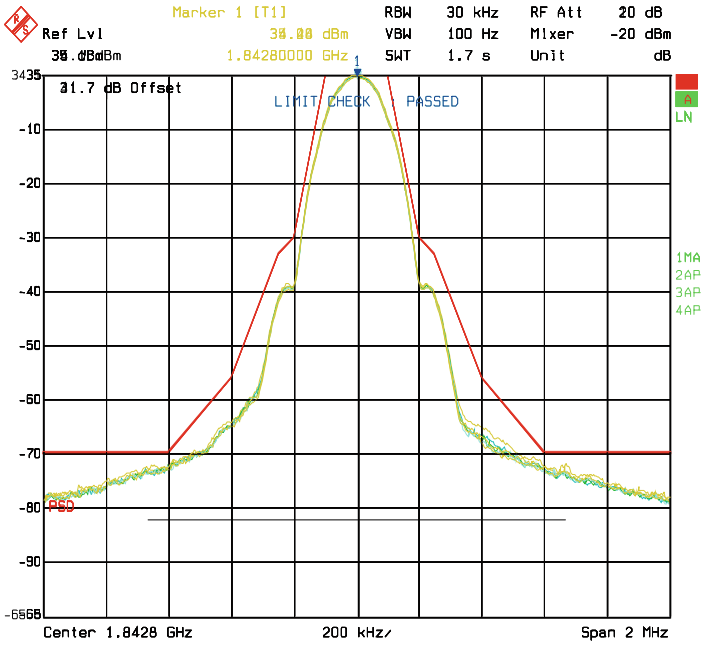


Fig. 13.11 Overlay of nine output spectral quality measurements for EDGE, with temperature swept from 30°C to 70°C and back again. Almost no changes are seen in both power level and signal spectrum, without any thermal compensation

condition shown in Fig. 13.4 calling for the switch ON resistance to be negligibly small compared to the stage load resistance/impedance. When this condition is met, if the actual ON resistance varies with temperature (and/or with manufacturing variances) then there is still a very small change in actual current flowing through the load. This means that the output signal is similarly stable. Figure 13.11 shows that not only is the signal power very stable across temperature sweeps, but signal quality also is extremely stable. This stability is inherent to this approach, so there is no need for additional thermal compensation circuitry.

13.5 Conclusions

By inverting the conventional design process for multimode transmitters, it is possible to realize energy efficiencies that are much higher than correspondingly provided by linear design approaches. At the same time, output signal quality from this switch based architecture is very good, showing large margins to the required specifications. Multimode operation is checked for GSM, EDGE, UMTS, cdma2000, Bluetooth, and OFDM. Even using the lowest efficiency mode of polar operation, the measured energy efficiency is higher than the best reported to date from linear or

linearized transmitters, particularly for these signal quality levels and output power values. Further improvements are expected as switching techniques are added to the envelope modulation block.

References

1. E. McCune, “Gain for Compressed Amplifiers”, Microwave Journal, May 2002.
2. W. Sander, et.al., “Polar Modulator for Multi-Mode Cell Phones”, Proceedings of the Custom Integrated Circuits Conference, IEEE, May 2003, pp. 439–445.
3. R. Pullela, et.al., “An Integrated Closed-Loop Polar Transmitter with Saturation Prevention and Low-IF Receiver for Quad-band GPRS/EDGE”, Proceedings of the International Solid State Circuits Conference, IEEE, February 2009, pp. 112–113.
4. E. McCune, “High-Efficiency Multi-mode, Multi-band Terminal Power Amplifiers”, IEEE Microwave Magazine, March 2005, vol. 6, pp. 44–55.

Chapter 14
RBS High Efficiency Power Amplifier Research – Challenges and Possibilities

Bo Berglund, Ulf Gustavsson, Johan Thorebäck, and Thomas Lejon

Abstract This text gives an overview of critical technology challenges for high efficiency power amplifiers used in future mobile broadband systems. Implementation aspects of efficient wideband multi-band transmitters are discussed. Simulations of and design implications for a multi-band 1.8–2.7 GHz high efficiency power amplifier using GaN transistors are presented. Finally possible transmitter architectures with potential of meeting ambitious efficiency, flexibility and frequency range goals are briefly analyzed.

14.1 Introduction

To increase radio base station (RBS) power amplifier (PA) efficiency and linearity has been in focus for radio technology research for many years. High efficiency Doherty RF power amplifiers combined with digital pre-distortion (DPD) linearization is now established as the primary technology of choice in demanding mobile broadband applications. Drain efficiencies over 50% for WCDMA signals have been reported for Gallium Nitride (GaN) Doherty amplifiers combined with DPD in the recent years, [1].

A major trend in the mobile systems industry is the push for higher data rates whilst maintaining high spectral efficiency. This is achieved by using higher order modulation like 16-QAM and 64-QAM as well as implementation of multiple-input multiple-output (MIMO) technology for certain radio scenarios. In parallel, wider operating bandwidths are used to increase data rates even further. Bandwidths are increasing from typically 5 up to 20 MHz, and future bandwidths up to 40 MHz can be anticipated. Good system signal to noise ratio (SNR) is required to reap the benefits of these measures. As a consequence increased signal accuracy, measured as error vector magnitude (EVM), is required from the RBS transmitter.

B. Berglund (✉), U. Gustavsson, J. Thorebäck, and T. Lejon
Ericsson AB, Kista, Stockholm, Sweden
e-mail: bo.g.berglund@ericsson.com

A.H.M. van Roermund et al. (eds.), *Analog Circuit Design: Smart Data Converters, Filters on Chip, Multimode Transmitters*, DOI 10.1007/978-90-481-3083-2_14,
© Springer Science+Business Media B.V. 2010

To fulfill the goals of higher data throughput over wide areas, higher average output power is requested. Simply, by going from 5 MHz carrier bandwidth up to 20 MHz, the transmitted output power needs to increase by a factor of four in order to maintain the coverage area. To avoid extremely high average powers a real deployment will most likely use a combination of smaller cells, high antenna gain and higher power levels. This can be anticipated as higher power levels also creates higher interference levels.

A thorough discussion of the 3GPP evolution can be found in [2]. The mobile broadband evolution thus drives the need of power amplifiers with higher average output power and higher linearity. This has to be achieved in a context of shrinking RBS physical footprint at the same time as the number of radio chains increases with the introduction of MIMO technology.

The energy cost for operating a radio access network, RAN, has to come down and the environmental impact of the radio networks must be minimized. The efficiency of the RBS power amplifiers have to improve as they are the major consumers of energy in an RBS. A detailed overview of RAN environmental impact is given in [3].

Expansion of mobile broadband data traffic also drives the need for wider spectrum usage i.e. usage of a wider range of frequency bands. In the latest 3GPP UTRA FDD RBS specification 14 frequency bands ranging from 698–2690 MHz have been identified for 3G operation; frequency bands up to at least 4200 MHz are candidates for mobile broadband application. The number of spectrum arrangements accelerates with new bands, but also the variation within a band between regions. Clearly, multi-band power amplifier technology is in demand to enable frequency flexible radio transmitters.

14.2 Power Amplifier Efficiency in an RBS Transmitter Context

The theoretical drain efficiency of a Class B amplifier is proportional to the square-root of the RF output power. Hence, the drain efficiency will decrease as the instantaneous output power is reduced from the maximum peak power of the amplifier [4].

The amplifier efficiency drops drastically in back-off as indicated in Fig. 14.1. This is a significant problem for RBS transmitters that operate with signals having large envelope variations. Peak-to-average ratio (PAR) is used as a measure to quantify the signals envelope variations. HSPA and LTE modulation have practical PAR values of approximately 10 dB. By using PAR reduction schemes, like clipping, the actual PAR can be reduced from 10 dB down to the 6–8 dB range, the actual value will depend on EVM and spectrum purity requirements. Clipping will however increase the signal EVM and is thus clearly a limiting factor when using higher order modulation, [5]. To provide low EVM in future high order MIMO deployments RBS transmitters should preferably be operated at fairly high PAR values around 8–9 dB.

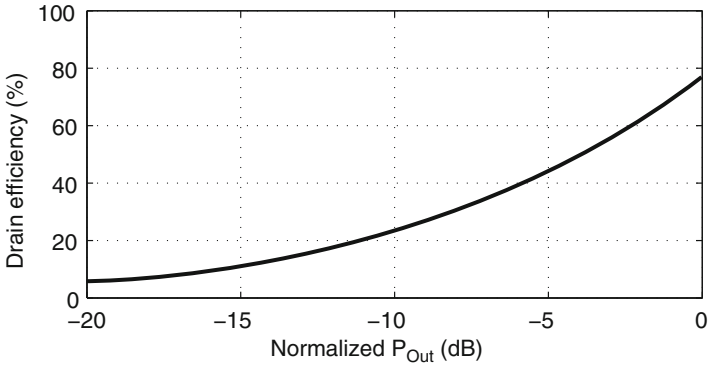


Fig. 14.1 Theoretical Power Efficiency of a Class B Amplifier

When deployed in a real network the RBS PA is typically backed-off even further to create network planning margins. The RBS should operate at full average power only in a fully loaded network scenario. Typically, measured over time, the PA will operate 3–4 dB backed-off from nominal average power; still high efficiency is requested from an energy saving perspective. In reality a good PA architecture should provide high efficiency over a 11–15 dB power range from saturated power. The theoretical Class B efficiency has dropped from 78% to 15% at 15 dB back-off as indicated in Fig. 14.1. Thus, there is a demand for a wideband RBS transmitter architecture that can provide high efficiency even at deep back-off.

14.3 Software Defined Radio (SDR) RBS

14.3.1 SDR RBS

To meet the needs of mobile broadband evolution and the associated expansion of spectrum use more flexibility is required from future RBS products. Radio access technologies, RAT, are migrating towards systems with higher data rates like High-Speed Packet Access (HSPA) and Long-Term Evolution (LTE) at the same time as new frequency bands are taken into operation. For operators this means that there is a need for flexible systems that can adapt to new RAT as well as providing frequency reconfigurability and possibilities to adapt output power without sacrificing efficiency. SDR with flexible RF front-end is a way to meet these requirements. Reducing equipment physical footprint is critical to provide cost efficient RBS site solutions. MIMO introduction with the associated increase of radio chains thus creates a need of highly integrated transmitter architectures.

Operational features of a future SDR RBS solution can be summarized as:

- **Multi-standard operation, MSR:** A radio used for different RAT's however not necessarily at the same time.

- **Mixed-mode operation:** An MSR operated with at least two modulation schemes, e.g. GSM and LTE, simultaneously. This type of systems put very demanding requirements on the clipping and linearization schemes to maintain efficiency.
- **Frequency reconfigurability:** A radio that is reconfigurable over a wide frequency range. Operation can be adjusted to any specific operating bandwidth over a large frequency range.
- **Highly integrated RF architecture:** A radio supporting multi order MIMO and shrinking RBS physical footprint

The main obstacles that so far have hampered the development of a generic SDR have been the lack of wideband RF technology with a wide frequency range providing acceptable cost/performance trade-off.

14.4 Wide RF Bandwidth Power Amplifier Design

Wide band designs of the power amplifiers for cellular infrastructure have proved difficult because of large parasitic capacitors of the commonly used laterally diffused metal oxide semiconductors (LDMOS) RF power transistors. They create a low impedance environment at the interface of the die which is in the range of 1 Ω or lower. The necessary high Q impedance transformation up to 50 Ω is challenging for a wideband design.

A common way of designing a power amplifier for base stations is shown in Fig. 14.2. A shunt L inductance (L_3) is used in parallel with C_{ds} to create a high impedance resonator and a T-match (L_1, L_2 and C_1) is used at the gate to transform the low impedance set by C_{gs} to a higher value. Wide bandwidth is achieved by mounting the resonators close to the transistor die, often in the package where the transistor die is mounted. The task to match the packaged device to a 50 Ω interface is solved on a Printed Circuited Board (PCB).

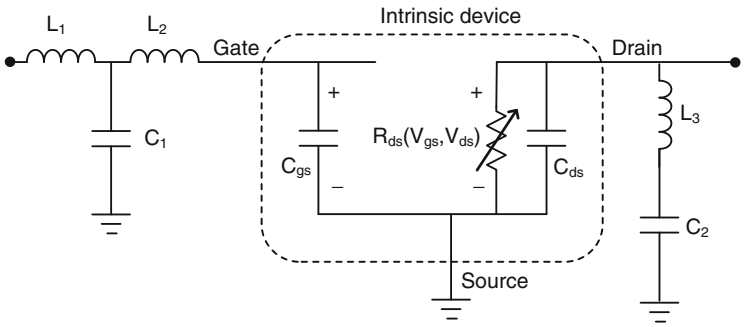


Fig. 14.2 A typical, yet simplified, radio base station power amplifier implementation

Table 14.1 C_{ds} for 50 W devices of two common technologies, [6]

Technology (Nominal V_{dd}, V)	C_{ds} (pF)	Optimum load impedance (R_{opt}, Ω)
28 V LDMOS	18	7.8
48 V GaN HEMT	2.3	23

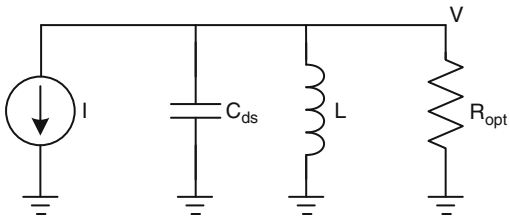


Fig. 14.3 A simplified equivalent circuit for the output network of a power amplifier

Recent development of low parasitic GaN RF power transistors opens up possibilities for multi-band designs since the impedances at the die are large. This lowers the Q value of the impedance transformation networks and consequently increases useful bandwidth. As seen in Table 14.1 the output capacitance of a GaN is reduced by almost a decade compared to LDMOS.

Figure 14.3 shows a simplified output network of a Power Amplifier. The transistor is described as a current source which is true when the transistor works in its linear region, C_{ds} is the parasitic output capacitor of the transistor and L forms a parallel resonator together with C_{ds} resulting in a purely real load, R_{opt} seen by the current source at the resonance frequency.

$$\frac{V}{I} = -\frac{jR_{opt}\omega L}{R_{opt}(1-\omega^2 LC_{ds}) + j\omega L} \tag{14.1}$$

Equation 14.1 is the analytical transfer function of the network in Fig. 14.3. Figure 14.5 show frequency sweeps of 14.1 using the values in Table 14.1. With GaN the frequency range from 1.8 GHz to 2.7 GHz can be covered with only a few tenths of dB in loss. Figure 14.4 show a block diagram of a simulated example of a wideband power amplifier design using a commercial GaN transistor model.

Transforming the impedance up from R_{opt} to 50 Ω is usually implemented by one or more low pass LC networks, the number of necessary stages for a wide band design is set by the impedance transformation ratio and the need to keep the total network Q value low. GaN with its larger R_{opt} is easier to transform, since the transformation ratio is lower and thus fewer stages are required.

The bandwidth limiting part in the design is the input match where the gate to source capacitor is large. A two stage LC matching network is used to transform the input impedance up to 50 Ω and bandwidth is traded for return loss in the optimization of the network. The poor input return loss is then handled by connecting the amplifiers in hybrid coupling. A shunt L inductance and one LC stage for the impedance transformation up to 50 Ω is used on the output according to the analysis

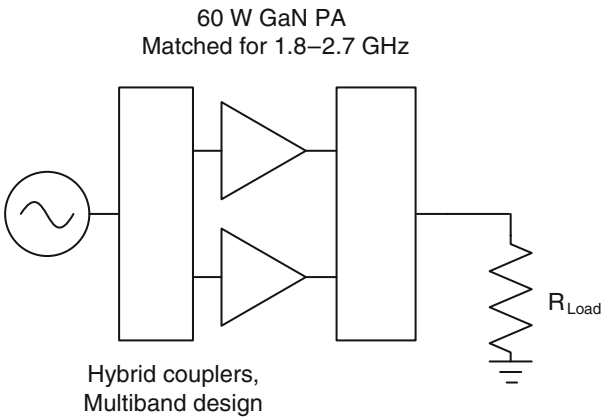


Fig. 14.4 Simulated hybrid PA setup with two 60 W wideband GaN power amplifiers and hybrid couplers designed for multiband operation

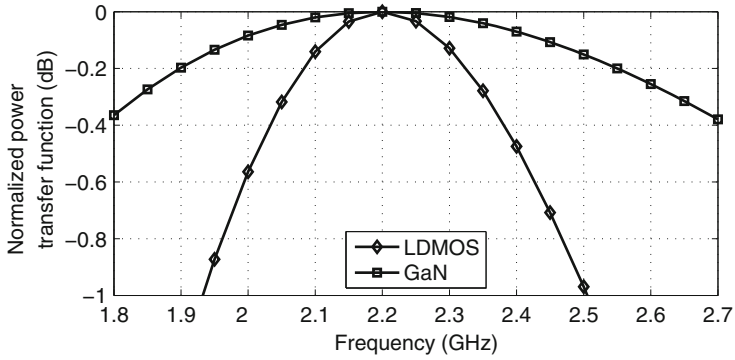


Fig. 14.5 Bandwidth comparison of output networks with LDMOS and GaN

above. Figure 14.6 shows a simulated frequency sweep of output power and power added efficiency showing more than 40% of relative bandwidth which covers several major cellular frequency bands between 1.8 to 2.7 GHz.

14.5 Efficiency Enhancement Techniques

An efficient RBS transmitter requires power amplifier architectures that provide high efficiency even at deep back-off levels. A classical Doherty implementation, [4], gives high efficiency at 6 dB back-off but efficiency drops rapidly at higher back-off values. The Doherty output network is also band limited as it uses a quarter wave transmission lines as impedance inverter. Methods like asymmetrical Doherty and multi-order Doherty can be utilized to provide high efficiencies at higher back-off values.

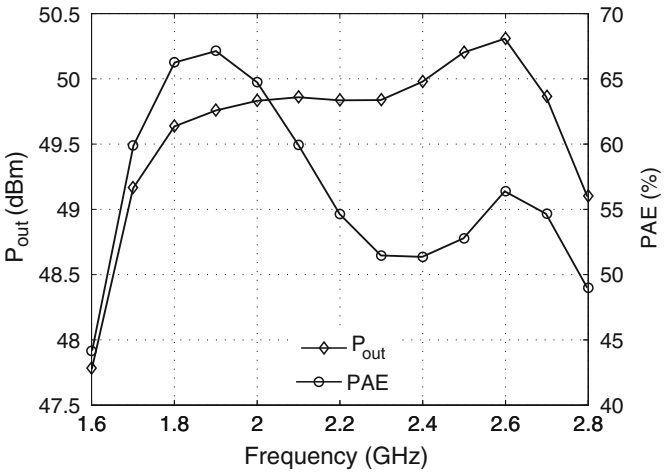


Fig. 14.6 PAE and output power vs frequency of a multi-band PA design using GaN transistors

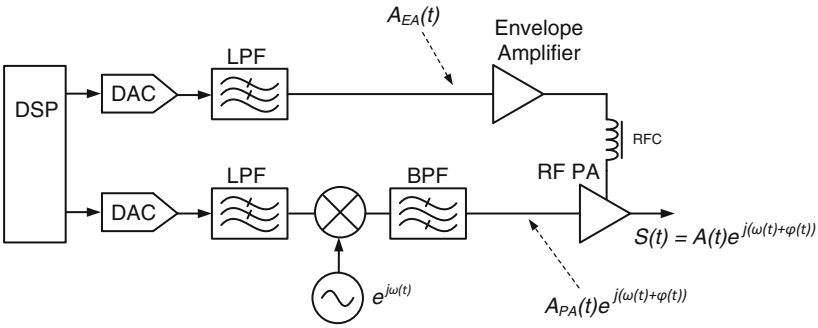


Fig. 14.7 An overview of a Envelope-Tracking system. $A_{EA}(t)$ is the amplitude-dependent signal sent to the envelope amplifier in order to track the envelope of the signal. $A_{PA}(t)e^{j(\omega(t)+\varphi(t))}$ is the amplitude and phase modulated carrier applied to the power amplifier

Envelope tracking, see Fig. 14.7, is an interesting alternative for multi-band amplifiers since the efficiency enhancement function is not dependent on the RF signal phase as in Doherty but instead works by changing the drain bias voltage with the envelope of the signal. The multi-band high efficiency design task is translated to the design of a wideband RF choke.

Figure 14.8 shows a simulation on possible efficiency under Envelope tracking on the design in Fig. 14.4, where the main improvement over a Class AB PA is observed at moderate back-off levels around 10 dB. The efficiency varies with frequency, mainly because of varying harmonic impedances seen by the transistor over the frequency range, and the research challenge is to design a multi-band amplifier with equal and high efficiency on all bands.

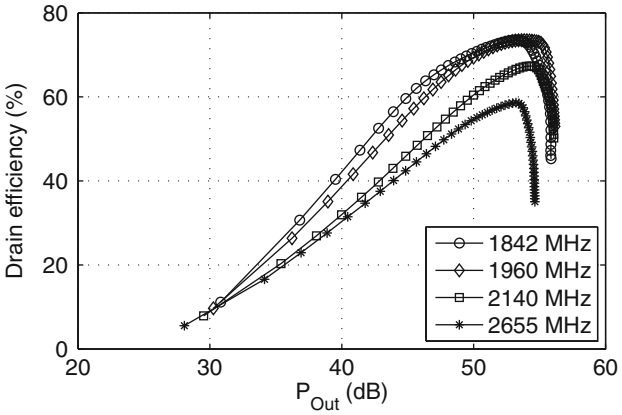


Fig. 14.8 Envelope tracking efficiency of the multi-band PA design in Fig. 14.4

14.5.1 Envelope Tracking Transmitter Architecture

From the expression of drain-efficiency for a Class B amplifier [4], one can note the inverse relation between efficiency and drain-supply voltage.

$$\eta = \frac{P_L}{P_{DC}} = \frac{I_0^2 \cdot R}{8} / \frac{V_{dd} \cdot I_0}{\pi} = \frac{\pi R I_0}{8 V_{dd}} \tag{14.2}$$

As long as the drain-supply voltage is high enough to keep the PA out of saturation, one can lower the drain-supply voltage and improve efficiency. The concept of Envelope Tracking is simple and intuitive; by following the RF envelope, keep the drain supply voltage to a minimum, without saturating the transitor.

The design challenge for an Envelope Tracking transmitter is to design the power supply to the amplifier. It should provide a variable output voltage with low output impedance over the baseband bandwidth of interest.

As the extracted amplitude of the RF envelope will be much wider than the original signal, the envelope amplifier has to be designed to handle wide bandwidth baseband signals, which is a challenging design task, bearing in mind the high efficiency requirements of the envelope amplifier.

The efficiency of the envelope tracking system is very dependent on the efficiency of the envelope amplifier, which most likely will be worse than the 90% efficiency of a typical fixed power supply.

The total efficiency of the envelop tracking system is the product of the envelope amplifier efficiency (η_{EA}) and the drain modulated PA's efficiency (η_{PA}), according to:

$$\eta_{\mathrm{TotET}} = \eta_{\mathrm{EA}} \cdot \eta_{\mathrm{PA}} \tag{14.3}$$

The total efficiency of envelope tracking system should be compared to the normal Class AB PA with its fixed power supply, who's total efficiency is calculated in a similar manner in Eq. 14.4, where η_{DCDC} is the efficiency of the DC/DC supply and η_{AB} is the efficiency of a class AB PA.

$$\eta_{\mathrm{TotAB}} = \eta_{\mathrm{DCDC}} \cdot \eta_{\mathrm{AB}} \tag{14.4}$$

One way to design a wide-band and efficient envelope amplifier is to use an architecture with a switcher stage and a linear amplifier in parallel. The switcher stage will provide the low frequency bulk of the amplitude signal with high efficiency. The high frequency content will be provided from a wide bandwidth linear stage, at a lower efficiency, [7].

By limiting the signal excursions down to zero by applying a floor to the amplitude signal, two benefits will arise;

1. The bandwidth of the amplitude signal will be smaller.
2. Reversed polarity between gate and drain will be avoided.

On the RF transistor side of the system, the varying supply voltage will have different impacts. As the drain voltage is kept to a minimum, shunt losses of the RF device will be less critical. This is valid for LDMOS transistors, however LDMOS has traditionally shown an increasing drain to source capacitance, C_{ds}, which will counteract the benefit of the lowered drain voltage.

Low series loss from the transistor will become more important, as the current generated by the transistor is the same as in the Class AB case. Implementations with Gallium Arsenide High Voltage Heterojunction Bipolar Transistors (GaAs HVHBT) with inherent low series-loss, have shown good results.

14.5.2 Third Order Doherty

An extension of the Doherty concept to a third order design will extend the high efficiency region down to deeper backoff levels compared to the standard Doherty. The third transistor also gives headroom for higher peak power with maintained average power which is important for coverage.

Figure 14.9 shows simulated power efficiency of a third order Doherty PA design using commercial GaN transistor models and micro-strip layout models including loss. The used ratio of peak to main transistor size is 3:1, selection of the peak to main ratio is critical for the exact efficiency vs. power characteristics. The simulation results are encouraging and are in line with simulations and measurements presented in [1].

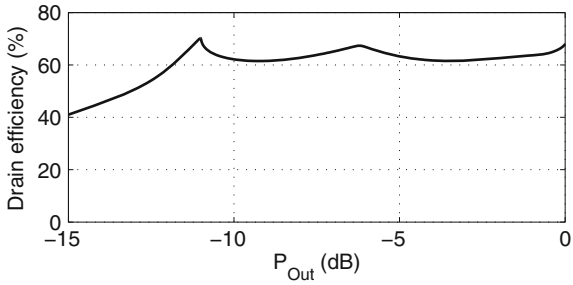


Fig. 14.9 Simulated efficiency of a third order Doherty

14.5.3 Pulsed Transmitter Architectures

Another method considered for improving the power efficiency in microwave power amplifier systems is to apply 1-bit quantization of different forms on the signal[1]. The key idea is to use the power amplifier only in its most efficient regions, i.e. in deep compression where the power efficiency is high or in off-state where the dissipated power is close to zero if operating near class C bias. Besides the prospect of high power efficiency, these types of architectures are usually considered for highly integrated systems as well.

One artifact of these pulsed modulation schemes is the distortion[2] produced when representing an amplitude-continuous signal with a binary signal. In order to comply with the system requirements certain measures has to be taken to ensure that the signal is properly reconstructed.

Two critical parameters are the pulse-rate and the output bandpass filter (or reconstruction filter). If the pulse-rate is to low, the filter needs a very narrow passband region in order to reconstruct the signal properly, as shown in Fig. 14.10. Such filters implemented at microwave frequencies usually have a insertion loss large enough to affect the power efficiency in a negative way.

On the other hand if the pulse-rate is increased, other parameters such as switch-loss and bandwidth limitations can be an issue. If the quantization scheme is implemented in a inefficient manner, the total power efficiency will also suffer due to low coding efficiency of the quantization as illustrated in [8].

Despite the limiting factors, we have seen some promising realizations of pulsed transmitter architectures in recent times, where either the PA input [9] or supply voltage [10] are pulsed for high efficiency operation.

In [11] it is suggested to use a feed forward method, a otherwise quite common method used to perform linearization to cancel out some of the quantization distortion around the passband of the filter. The feed-forward method does however

[1] This is hereafter denoted as pulsed modulation (PM).

[2] This distortion is generally referred to as *quantization noise*.

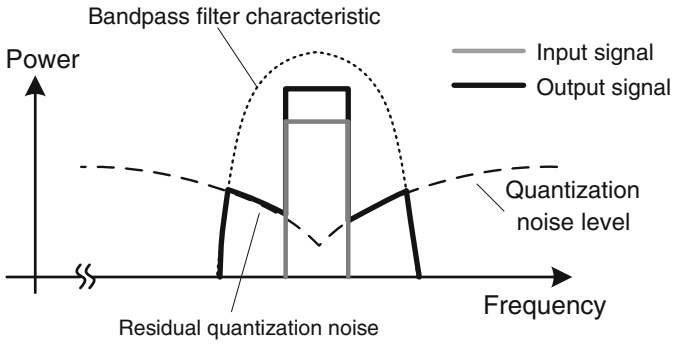


Fig. 14.10 An illustration of the quantization distortion and the bandpass-filter characteristics needed

require additional hardware which adds unnecessary cost and complexity to the system in the form of additional hardware which also might have a negative impact on the overall system efficiency and complexity. A low complexity solution is suggested in [12] that requires no need for additional hardware. All of these architectures performs quantization of the signal in some way. These methods of representing the amplitude-continuous signal in a pulse coded manner can be grouped into three main categories, namely;

- **PWM/PPM (Pulse Width and/or Position Modulation):** The amplitude and phase information of the communication signal is encoded by altering the width and position of the pulses respectively. Implementation issues regarding the possibility to encode signals with high dynamic range occur due to the amplitude mapping on the pulse-width. Very narrow pulses with a large harmonic content is needed to represent low amplitudes.
- **PDM (Pulse Density Modulation):** The amplitude and phase information of the communication signal is encoded by changing the density of the pulses. This type of encoding scheme is most common in oversampled modulator types, such as the $\Sigma\Delta$-modulator, [13]. PDM, and in particular the $\Sigma\Delta$-modulator, comes with a feature usually referred to as *noise-shaped coding* where the possibility of shaping the spectral contents of the quantization distortion is introduced, [14].
- **PAM (Pulse Amplitude Modulation):** The amplitude information of the communication signal is encoded by altering the amplitude of the pulses. Commonly used in communication systems where bandwidth and power efficiency is a lesser issue, such as fibre optic systems.

Usually, PAM is not considered in these types of architecture since it would operate the amplifier in sub-efficient regions. However, PAM relates quite closely to the method described in [12], in which a small controlled amplitude component is superimposed on the quantized signal, cancelling out a selected band-pass part of the quantization distortion close to the carrier. This enables the use of a more wide-band output filter with less insertion loss. Also, operating at quite moderate sampling rates, avoiding large switch losses, is possible.

14.5.3.1 Three Examples of Pulsed Transmitter Architectures

We will now review three examples on pulsed transmitter architectures that are commonly suggested in the literature where a communication signal

$$S[n] = \underbrace{A[n]e^{j\varphi[n]}}_{\text{Polar form}} = \underbrace{I[n] + jQ[n]}_{\text{Cartesian form}} \tag{14.5}$$

is quantized by different methods in either polar or Cartesian form.

Baseband or Envelope PM In general, when speaking of baseband or envelope PM, one is referring to the process of performing quantization on the amplitude component or the envelope ($A[n]$) of the communication signal, Eq. 14.5. This signal is then simply up-converted onto the RF-carrier as shown in Fig. 14.11, before being passed on to the remaining transmitter chain (pre-driver, driver, PA, filter, e.t.c).

Another possible solution would be to switch the drain-voltage supply, V_{dd}, with the pulse coded amplitude information as discussed in [15]. One advantage of baseband PM is that the quantization is performed on the baseband signal before being converted up to RF, thus unlike RF PM, the quantization scheme is completely independent of the carrier frequency.

Cartesian PM As suggested in [16], the quadrature components of the communication signal is quantized separately, as shown in Fig. 14.14. A time-interleaved combination is then performed, after which the quantized quadrature signal is up-converted to RF via a wide-band mixer. In comparison to baseband PM, Cartesian PM has the advantage of working on the I/Q signals, which in general are more band-limited as opposed to the envelope, $|S[n]|$, which has a much larger bandwidth.

Similar to baseband PM, Cartesian PM is done on baseband which makes the quantization scheme carrier frequency independent.

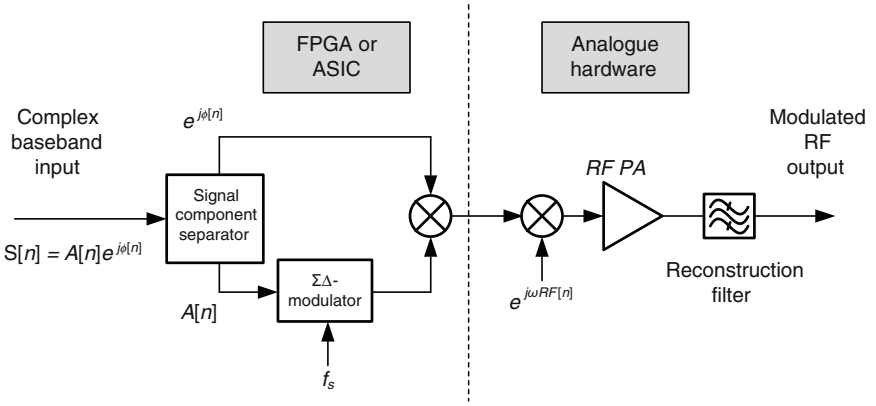


Fig. 14.11 Example of a baseband PM system

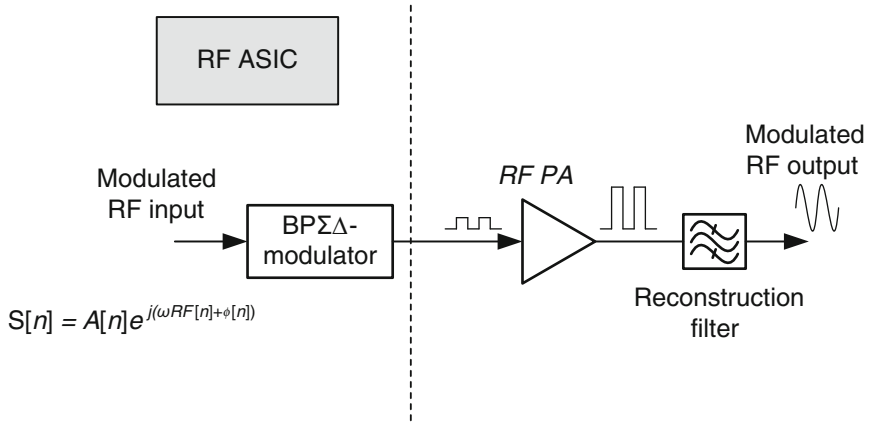


Fig. 14.12 Illustration of an RF PM system

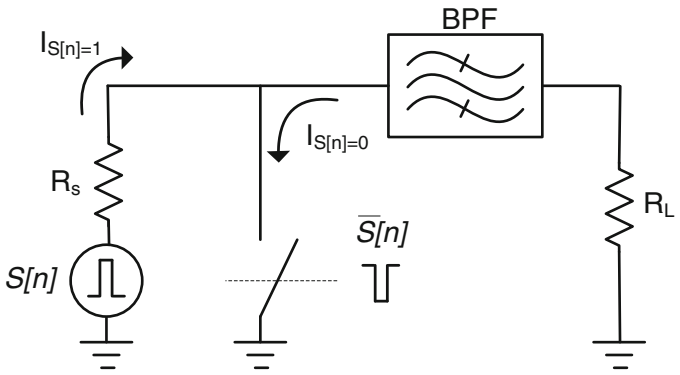


Fig. 14.13 A simplified sketch illustrating the switched resonator principle

RF PM As suggested in [8], a so called band-pass ΣΔ-modulator is used to perform quantization on the modulated carrier as shown in Fig. 14.13. PWM/PPM is also considerable as suggested in [17]. These types of architecture are usually considered for highly integrated systems due to its simplicity in terms of few building blocks.

One drawback of this method however, is that the quantization needs to be performed on at least a rate corresponding to the carrier frequency which leads to implementation issues in terms of bandwidth, in particular for the higher frequency bands. At these rates the gate-source capacitance of the power amplifier device, C_{gs}, in combination with the bond-wires used to connect the device to its environment forms a lowpass filter that might prevent the harmonic content of the pulse-train to be delivered to the gate properly, thus preventing the power amplifier to switch properly.

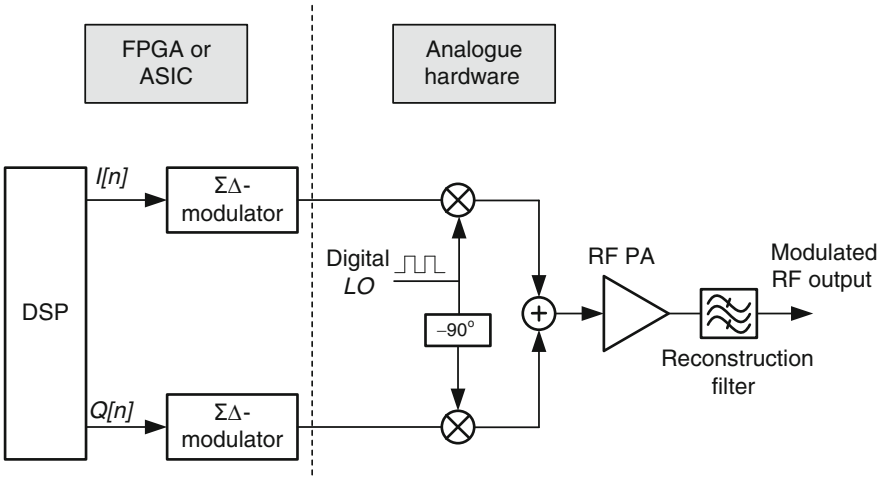


Fig. 14.14 Illustration of a Cartesian PM system

14.5.3.2 Efficient Filtering of the Residual Quantization Distortion

One issue still remaining is how to deal with the residual error in a good way–keeping the power efficiency high while not compromising the system requirements. One effect that should be taken in to consideration is the discharge of the remaining energy located in the reactive elements in the bandpass filter back into the device during its "off"-state.

One proposed method of dealing with this is to use a so called *switched resonator technique* where the residual energy is discharged to a ground plane instead of back in to the active device. This is depicted in Fig. 14.14 with a simplified sketch. It is shown in [18] that a regular second order Doherty amplifier inherently performs this discharge via the peak amplifier. The principle is then demonstrated on a second order, 1.8 GHz GaAs Doherty with promising results.

14.6 Conclusions

A number of demanding technical challenges must be overcome to establish a flexible high efficiency RBS transmitter architecture. The main objective will be to improve efficiency over a wide range of output power while at the same time improving linearity, increasing operating bandwidth and introduce spectrum flexibility.

Recent developments in high power transistor technology, like GaN, have drastically improved the possibilities for practical realization of wideband highly efficient RF power amplifiers. A combination of wideband power amplifiers and

efficiency enhancement technologies, like Doherty and Envelope Tracking, show good potential of meeting critical future RBS characteristics requirements.

Continued research is however needed on highly integrated flexible and efficient RBS transmitter architectures in particular for higher order MIMO application.

Acknowledgments The authors would like to thank Daniel Åkesson (Ericsson AB) for the third order Doherty simulations and Hossein Mashad Nemati, Mattias Thorsell and Thomas Eriksson at Chalmers University of Technology for proof-reading this text.

References

1. M.J. Pelk, W.C.E. Neo, J.R. Gajadharsing, R.S. Pengelly, and L.C.N. de Vreede, "A high-efficiency 100-W GaN three-way Doherty amplifier for base-station applications," *IEEE Transactions on Microwave Theory and Techniques*, vol. 56, pp. 1582–1591, July 2008.
2. E. Dahlman, S. Parkwall, J. Sköld, and P.Beming, *3G Evolution HSPA and LTE for Mobile Broadband*. Academic, ISBN 978-0-12-374538-5, 2008.
3. T. Edler and S. Lundberg, "Energy efficiency enhancements in radio access networks," *Ericsson Review*, vol. 81, no. 1, pp. 42–51, 2004.
4. S. Cripps, *RF Power Amplifiers for Wireless Communications*. London: Artech House, 2006.
5. O. Väänänen, J. Vankka, and K. Halonen, "Effect of baseband clipping in wideband cdma system," in *IEEE International Symposium on Spread Spectrum Techniques and Applications*, 2004.
6. W. Burger, "RF-LDMOS: an ideal device technology for ISM to WiMAX?" *Freescale Semiconductors, Presented at IMS*, Honolulu, Hawaii, 2007.
7. "Ericsson patent US6,3000,826 B1."
8. T. Johnson and S.P. Stapleton, "Comparison of bandpass $\Sigma\Delta$ modulator coding efficiency with a periodic signal model," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 55, no. 11, pp. 3763–3775, Dec. 2008.
9. C. Berland, I. Hibon, J. Bercher, M. Villegas, D. Belot, D. Pache, and V. Le Goascoz, "A transmitter architecture for nonconstant envelope modulation," *Express Briefs, IEEE Transactions on: Circuits and Systems II*, vol. 53, no. 1, pp. 13–17, Jan. 2006.
10. J. Choi, J. Yim, J. Yang, J. Kim, J. Cha, D. Kang, D. Kim, and B. Kim, "A $\Delta\Sigma$ -digitized polar RF transmitter," *IEEE Transactions on, Microwave Theory and Techniques*, vol. 55, no. 12, pp. 2679–2690, Dec. 2007.
11. T. Matsuura and H. Adachi, "A high efficiency transmitter with a Delta-Sigma modulator and a noise cancellation circuit," in *European conference on wireless technology*, Amsterdam, 2004.
12. U. Gustavsson, T. Eriksson, and C. Fager, "A general method for quantization noise suppression in pulsed transmitter architectures," *Accepted to IMS, Boston*, 2009.
13. N. Jayant and P. Noll, *Digital coding of waveforms*. Prentice Hall, Signal Processing Series, ISBN 0-13-211913-7, 1984.
14. R. Schreier, "Noise-shaped coding," Canada, 1991.
15. M. Nielsen and T. Larsen, "A transmitter architecture based on delta–sigma modulation and switch-mode power amplification," *IEEE Transactions on Circuits and Systems*, vol. 54, no. 8, pp. 735–739, Aug. 2007.
16. P. Kenington, *RF and Baseband Techniques for Software Defined Radio*. Boston: Artech House, 2005.
17. M. Nielsen and T. Larsen, "An RF pulse width modulator for switch-mode power amplification of varying envelope signals," *Topical Meeting on Silicon Monolithic Integrated Circuits in RF Systems*, pp. 277–280, 10–12 Jan. 2007.
18. J. Jeong and Y. Wang, "Envelope switched Doherty power amplifier for RF applications," *SBIR Phase II report*, 2005.

Chapter 15
Multi-Mode Transmitters in CMOS

Manel Collados, Xin He, Jan van Sinderen, and Raf Roovers

Abstract This paper both describes a multi-mode modulator for cellular 2.5, 3 and 4G applications and a multi-mode transmitter for connectivity standards like Bluetooth (BT), Zigbee and WLAN. Both multi-mode transmitters are implemented in deep submicron CMOS technology and demonstrate the use of digital techniques and innovative analogue circuit topologies for obtaining high efficiency and excellent performance. The two multi-mode transmitters have different architectures as their target applications impose different performance requirements.

15.1 Introduction

The need for new wireless services and higher data throughput are driving the creation of new wireless standards. In general these new wireless systems use more spectrally efficient modulation schemes and occupy more bandwidth. Common for these modulation techniques is the need for linear transmitters to preserve the desired signal properties like occupied bandwidth and non-constant envelope.

Multiple wireless standards for identical applications are coexisting and worldwide coverage is desired. As a result there is a clear trend for multi-mode and multiband transceiver solutions. This has triggered the concept of software defined radio (SDR). According to this principle, it is desirable to design radios which can cover as many standards as possible provided that performance is not compromised.

Another trend in mobile communications is the widespread use of CMOS technology replacing bipolar and BiCMOS solutions. This allows the integration of the radio together with the baseband on a single die, which reduces system costs. In order to extract the best radio performance out of the CMOS technology, both new CMOS analogue circuit topologies and an increase in the use of digital techniques [1] can be observed. These new techniques also help to reduce costs as they can lower the used silicon area, lower the required test-time and improve yield.

M. Collados (✉), X. He, J. van Sinderen, and R. Roovers
NXP Semiconductors, Eindhoven, The Netherlands
e-mail: manel.collados@nxp.com

A.H.M. van Roermund et al. (eds.), *Analog Circuit Design: Smart Data Converters, Filters on Chip, Multimode Transmitters*, DOI 10.1007/978-90-481-3083-2_15,
© Springer Science+Business Media B.V. 2010

Finally, battery life is a strong differentiator for mobile equipment. This fuels the quest for better overall transmitter efficiency. Unfortunately modern modulation techniques like OFDM have a relative high Peak-to-Average Power Ratio (PAPR) which prevents the use of efficient non-linear power amplifiers. Polar modulation techniques are theoretically very attractive to realize efficient transmitters for signals with non-constant envelope and high PAPR. Also new circuit topologies can improve efficiency significantly.

In FDD cellular systems such as WCDMA (3G) and LTE (4G), Tx and Rx operate simultaneously while only the duplexer provides isolation between them. Furthermore coexistence requirements imposed by other wireless systems that can be present in the same handheld device give very tough transmitter noise and spur requirements. Although an additional SAW filter between the modulator output and the PA could relax these requirements, this should be avoided as it increases system costs. In this paper a Cartesian architecture has been chosen for the cellular multi-mode modulator, as it can provide a very clean output spectrum with low noise and low spur levels. A novel direct quadrature voltage modulator using a passive voltage mixer driven by a 25% duty-cycle LO signal has been used to improve noise performance and efficiency.

In TDD connectivity systems such as Bluetooth (BT) and WLAN, Tx and Rx are operating in turn. Also emission mask requirements are relaxed compared to FDD cellular systems. Especially the high PAPR of the OFDM modulation used for WLAN makes the use of polar modulation very attractive for obtaining high efficiency. Practical implementations using PA supply modulation fall short in the bandwidth which they can efficiently achieve [2]. The connectivity multi-mode transmitter presented in this paper uses a digital polar concept which does not have these bandwidth limitations.

15.2 Direct Quadrature Voltage Modulator for Cellular Applications

As shown in Fig. 15.1, in FDD systems such as WCDMA and LTE, TX and RX operate simultaneously, while the duplexer provides the necessary isolation between them. In order not to desensitize the RX path, the TX noise in WCDMA band 1 has to be lower than $-155\,\mathrm{dBc/Hz}$ at 190 MHz offset (RX band), given a typical 45 dB TX-to-RX isolation from the duplexer [3]. Furthermore, when transmitting at 1,920 MHz, the nearest frequency offset to the DCS band (1,805 to 1,880 MHz) is 40 MHz. At this offset typical WCDMA duplexers only offer a few dB's attenuation. To comply with the 3GPP emission mask that restricts the emitted power below $-71\,\mathrm{dBm/100\,kHz}$ at the DCS band, the TX noise at 40 MHz offset is expected to be lower than $-148\,\mathrm{dBc/Hz}$ (with 3 dB margin). Nowadays there is a growing demand to integrate WCDMA and GPS applications in the same mobile device. In order not to interfere the GPS receiver, the far-out noise floor of the WCDMA TX is also desired to be very low. The overall noise floor requirement is shown in Fig. 15.2.

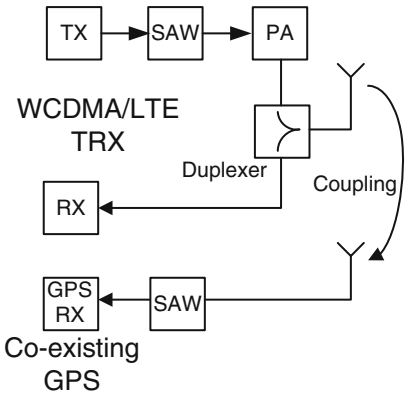


Fig. 15.1 WCDMA/LTE transceiver and co-existing GPS application in the mobile handset

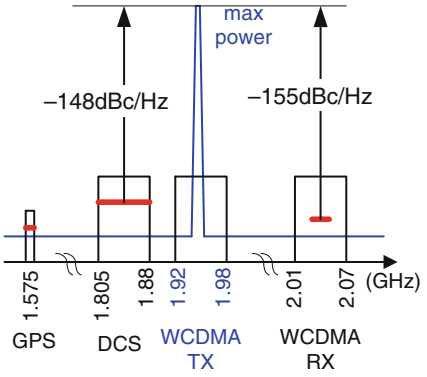


Fig. 15.2 Noise floor requirement for the WCDMA transmitter

Figure 15.3 shows the conventional Cartesian transmitter using active current (Gilbert) mixers. In practice the TX noise contributed by the PA is negligible. However, in the Cartesian transmitter the double-balanced current (Gilbert) mixer generates significant noise that necessitates a TX SAW filter between the transmitter and the PA. Targeting to eliminate the TX SAW filter, a current-quenching technique was proposed in [3] to lower the mixer noise, and a power mixer which delivers the required output power without the following PA driver was presented in [4]. Both approaches pay the price of high power consumption. Recently a low-power solution by introducing an integrated notch at the RX band was demonstrated in [5]. However, it is difficult to meet the noise requirement at the DCS band, and also the noise requirement for the co-existence of WCDMA and GPS applications.

15.2.1 Direct Quadrature Voltage Modulator

To address the low-noise low-power requirement of the WCDMA transmitter, this paper proposes an innovative approach using direct quadrature voltage modulation via a passive voltage mixer driven by 25%-duty-cycle LO [6]. As shown in Fig. 15.4,

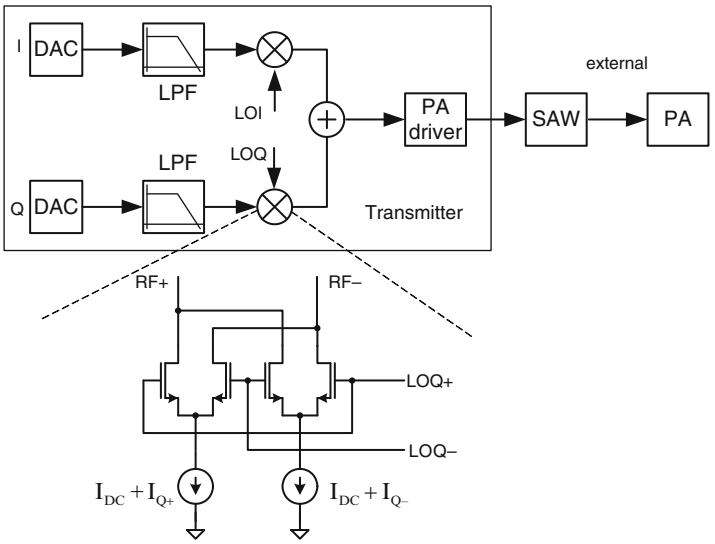


Fig. 15.3 The simplified diagram of the Cartesian transmitter using an active double-balanced current mixer (Gilbert mixer)

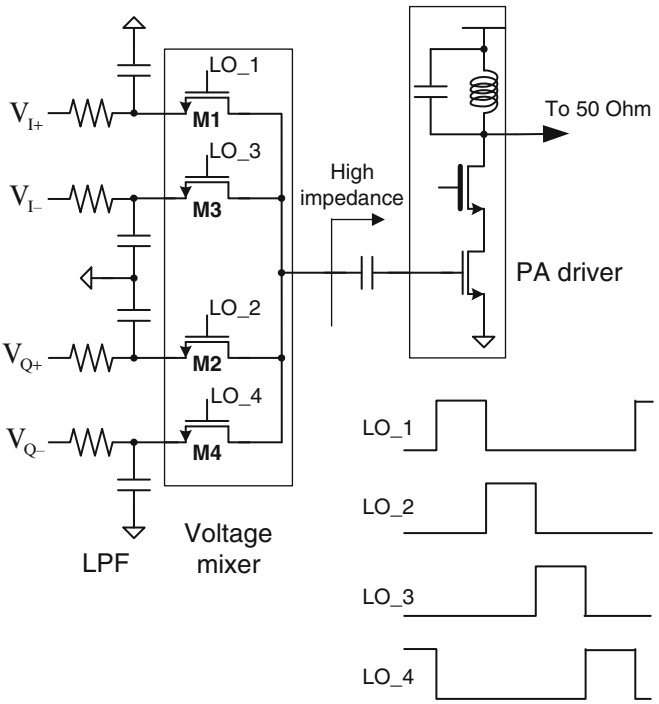


Fig. 15.4 Direct quadrature voltage modulator

first the IF I/Q voltage inputs are filtered by the passive LPF to lower the far-out noise. By switching on/off the transistors M1 to M4 through the quadrature-phased LO with 25% duty cycle, the filtered IF quadrature input voltages V_{I+}, V_{Q+}, V_{I-}, and V_{Q-} are sequentially copied to the voltage mixer output, where it sees the high input impedance of the PA driver. Such operation leads to a direct quadrature voltage modulation. The output of the voltage mixer contains both the desired fundamental LO mixing product, and the unwanted odd-harmonic LO mixing products. The conversion gain at the fundamental frequency for 25%-duty-cycle LO is about −0.9 dB, while the conversion gain at third order harmonic is roughly 10 dB lower. The unwanted harmonics can be attenuated below −40 dBc by a harmonic filter i.e., formed by the RF output bonding wire and capacitors.

In the voltage mixer only one switch is conducting at any time. Due to the high impedance presented at the mixer output, ideally there is no current flowing in the switches. However, considering the transition when one switch (e.g. M2) is turned on and the previous one (e.g. M1) is turned off, different voltage levels are presented at the drain and the source of M2, incurring current spikes that degrade the linearity. Fortunately the capacitors in the LPF can absorb those RF current spikes, and hence reduce the voltage disturbance to a very low level within several picoseconds.

To further improve the power efficiency, a class-AB PA driver is adopted. The PA driver employs a cascade structure which is powered at 1.8 V. For reliability a thick-oxide transistor is used for the cascade transistor.

The proposed approach achieves significantly improved performance over the conventional transmitter using Gilbert mixers. In the Gilbert mixer the voltage to current (V-I) conversion not only degrades the linearity, but also yields significant noise. To lower the noise to the required level, normally tens of mA currents have to be consumed. In addition, the V-I conversion mismatch between I/Q channels introduces an image, which degrades EVM. I/Q DC current offset also causes LO leakage to RF, which again degrades EVM at low output level. Using the proposed direct quadrature voltage modulation avoids those problems. The only noise generated by the voltage mixer is the thermal noise of the on-resistance when switches are turned on. With the scaling-down of CMOS processes, the size of the switch transistors can be further shrunk, resulting in reduced noise as well as power consumption in the LO generation circuit. Moreover, without V-I conversion the voltage mixer introduces much less I/Q mismatch, leading to improved EVM. LO leakage is also reduced, which is only caused by the capacitive mismatch.

15.2.2 25% Duty-Cycle LO Generation

By proper dimensioning the switch transistors, the noise generated in the voltage mixer is negligible. Now the far-out noise of the transmitter is mainly contributed by the LO phase noise. Aiming for low noise and low power, the implementation of the 25%-duty-cycle LO generation circuit is shown in Fig. 15.5. In the frequency divider, ideally the rising edges of the divider output LO_I+ and LO_I− coincide

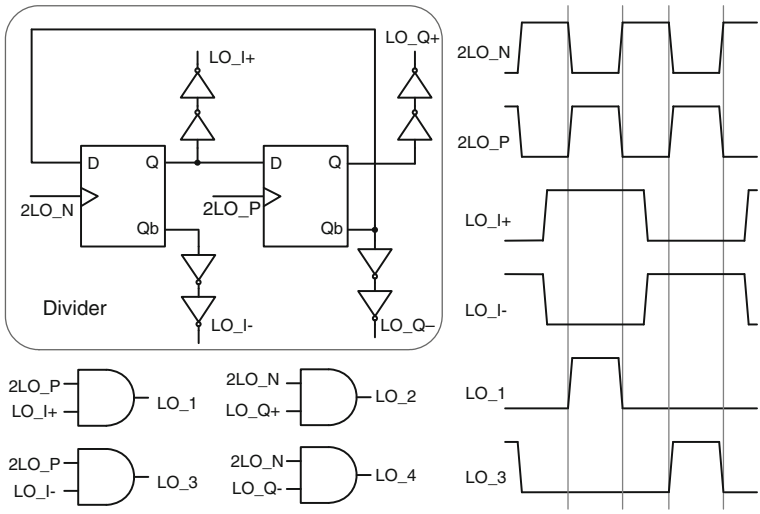


Fig. 15.5 Low-noise low-power 25%-duty-cycle LO generation circuit and corresponding waveforms

with the rising edge of the input clock 2LO_N, while the rising edges of LO_Q+ and LO_Q− coincide with the rising edge of 2LO_P. In the circuit two additional inverter stages are added to each quadrature output. The resulting delay, together with the settling time in the latches, shifts LO_I+ and LOI− to a phase where the pulse of 2LO_P falls inside the pulse of LO_I+, and the following pulse of 2LO_P falls inside the pulse of LO_I−. Hence the desired signals LO_1 and LO_3 are obtained by applying an AND function to LO_I+ and 2LO_P, LO_I− and 2LO_P, respectively. Similarly, LO_2 and LO_4 are obtained by applying an AND function to LO_Q+ and 2LO_N, LO_Q− and 2LO_N, respectively. In such configuration the rising edges and the falling edges of LO_1 to LO_4 are all derived from the input clocks running at twice the carrier frequency. Therefore the noise generated in the divider is not presented in the 25%-duty-cycle LO output, leading to a low-noise low-power implementation.

15.2.3 Measurement Results

Figure 15.6 shows the transmitter prototype taped-out. It includes the TX chain without DAC. The external input clocks runs at twice the carrier frequency, loaded by the buffer circuit on chip. A divide-by-2 is included in the 25%-duty-cycle LO generation circuit. The prototype is fabricated in a 45 nm CMOS process. As shown is Fig. 15.7. The die only occupies 0.8 mm^2, and the size of the active circuits is negligible.

Delivering 1 dBm WCDMA output at 1,950 MHz, the transmitter achieves 0.97% EVM and −49 dBc LO leakage, while consuming 9 mA at 1.8 V supply in the PA

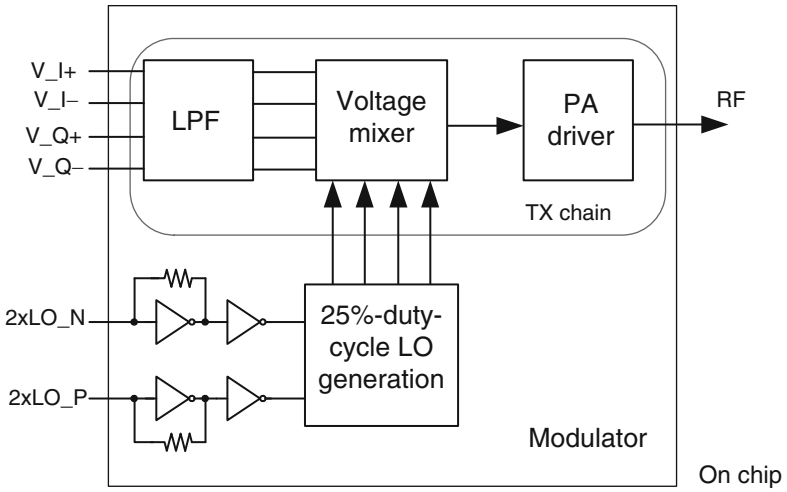


Fig. 15.6 Prototype of the direct quadrature modulator

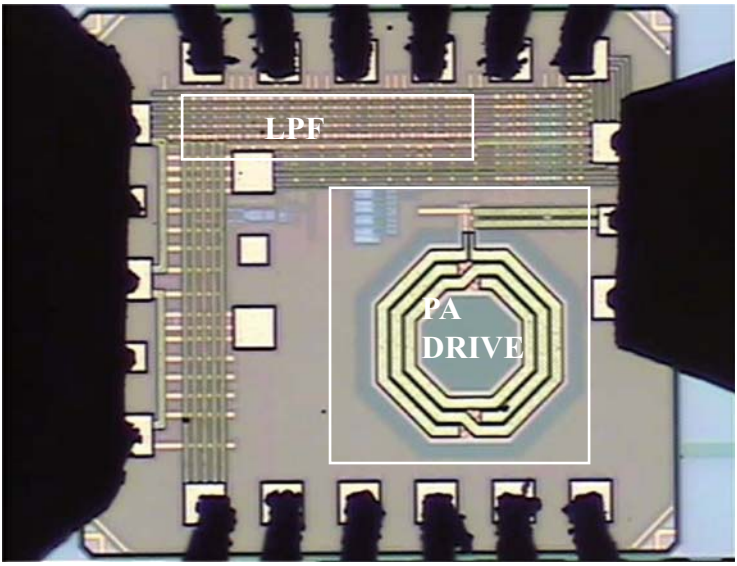


Fig. 15.7 Die micrograph

driver, and 5 mA at 1.1 V supply in the on-chip LO generation circuit (including buffers). Figure 15.8 plots the measured 1 dBm spectrum over 25 MHz, demonstrating −52 dBc ACLR at 5 MHz offset and −74 dBc ACLR at 10 MHz offset. The spectrum over 200 MHz is presented in Fig. 15.9, where the equipment noise floor is −118 dBm/10 kHz. The observed images at ±35 MHz offset are caused by the spurs from the external clock input. The noise at 1 dBm output is −159 dBc/Hz at

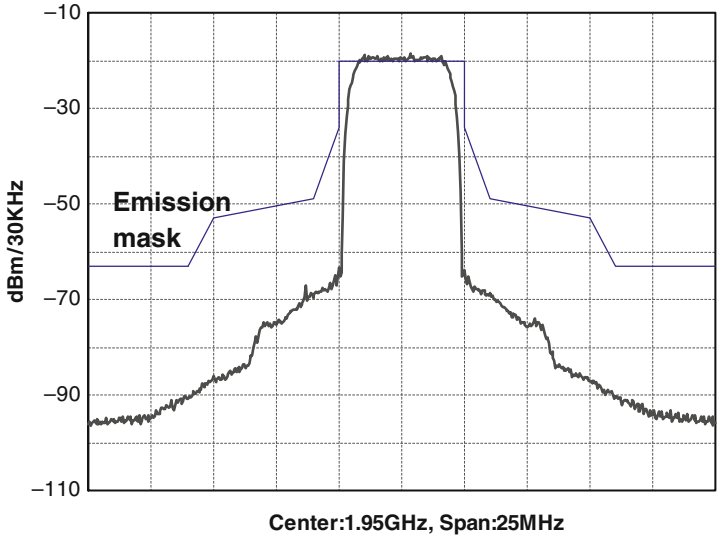


Fig. 15.8 Measured 1 dBm WCDMA output spectrum over 25 MHz range

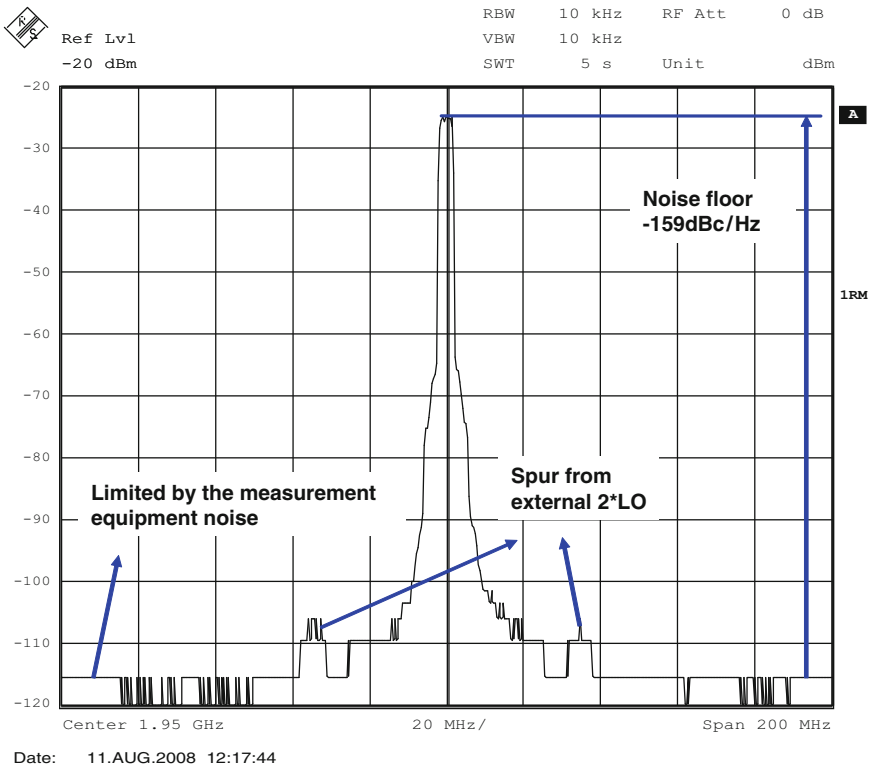


Fig. 15.9 Measured 1 dBm WCDMA output spectrum over 200 MHz frequency range

	Noise (dBc/Hz)	@ Offset (MHz)	EVM	ACLR @5/10 MHz (dBc)	Output power (dBm)	Power consumption (mW)	Power efficiency
[3]	–156	190	3%	–49 / –70	–10	113	0.09%
[4]	–158	40	3.7%	–46 / –72	3.8	235	1%
	–163	190					
[5]	–160 (notch)	190	4.5%	–44 / –58	3.3	65	3.3%
This work	–159	> 40	0.97%	–52 / –74	1	22	5.7%

Fig. 15.10 Performance comparison

the frequency offset beyond 40 MHz, sufficient to safeguard both the DCS band and the WCDMA RX band without needing a TX SAW filter.

Over 10 samples measured on the wafer, the best EVM is 0.88%, while the worst one is 1.1%. Notice the EVM of the arbitrary waveform generator used in the measurement is about 0.7%. After calibrating out the noise contributed from AWG, the final EVM is about 0.7%, which is much lower than the state-of-the-art.

Since the measurements have been performed on wafer directly without harmonic filter, the highest harmonic (third-order) level measured is about −20 dBc. Figure 15.10 summarizes the performance comparison. The improved performance of noise, ACLR, LO leakage, EVM, and power efficiency is achieved by using the proposed direct quadrature voltage modulation in the transmitter.

15.3 Digital Polar Transmitter for Connectivity Applications

Most of present day Zigbee, Bluetooth and WLAN solutions are based on the Cartesian decomposition. More recently, however, a number of transmitter architectures based on the polar representation of the baseband signal have been presented [1, 2]. At first glance, this does not seem a good idea. To begin with, extra digital signal processing is required to go from in-phase and quadrature samples to phase and envelope samples. Moreover, the new signals require more bandwidth, so higher sampling rates. Finally, the spectral purity of the transmitted signal relies on a good recombination of these two wide bandwidth signals, such that energy outside the wanted channel is cancelled out. So, why is the polar approach interesting? In the first place, it can help to improve efficiency. Cartesian transmitters require very linear PA's to preserve the input envelope modulation intact. This asks for the well-known back-off, with the consequent loss in efficiency. A polar decomposition of the baseband signal allows for more efficient topologies using voltage supply regulators, or dynamic biasing. Secondly, the use of the polar decomposition allows reusing existing transmitter building blocks like the frequency synthesizer or the power amplifier as modulators, reducing the overall number of transmitter blocks.

Also compared to a Cartesian transmitter, a polar transmitter does not suffer from I/Q imbalance or LO leakage due to DC offset in the I/Q chains. Finally, the fact that the VCO (or DCO) phase follows the transmitted signal phase makes it less susceptible to unwelcome pulling (coupling of transmitted signal power in the oscillator's LC tank).

In the following, we review two flavors of polar transmitter architectures before diving into an implementation of a multi-mode digital polar transmitter for connectivity.

In general, the transmitted RF signal of a wireless transmitter can be written in function of its quadrature components and the carrier frequency as

$$s(t) = A\left(x(t)\cos(w_c t) - y(t)\sin(w_c t)\right) \tag{15.1}$$

In a polar transmitter the envelope and phase signals are derived from the in-phase and quadrature signals as follows

$$r(t) = \sqrt{x^2(t) + y^2(t)} \tag{15.2}$$

$$\phi(t) = \arctan\left(\frac{y(t)}{x(t)}\right) \tag{15.3}$$

and the transmitted signal is obtained based on the equation

$$s(t) = A\left(r(t)\cos\left(w_c t + \phi(t)\right)\right) \tag{15.4}$$

A first transmitter architecture based on the polar signal decomposition is shown in Fig. 15.11. A similar type of transmitter is proposed in [7] for GSM/EDGE standards.

In this architecture, envelope samples are converted to the analog domain using a DAC followed by a reconstruction filter. The resulting analog signal is used to control an efficient DC-to-DC converter providing the supply voltage of a saturated PA. The assumption here is that the output RF envelope of the saturated PA is proportional to its supply voltage.

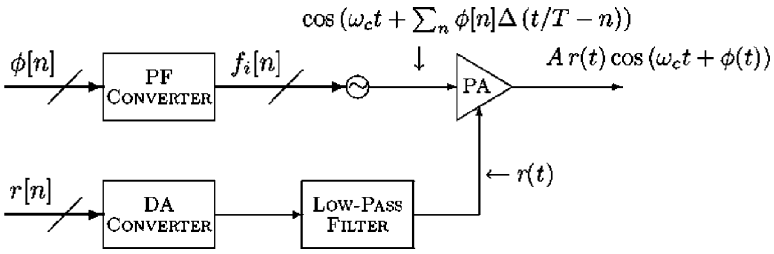


Fig. 15.11 Polar transmitter with digital frequency modulation and analog envelope modulation

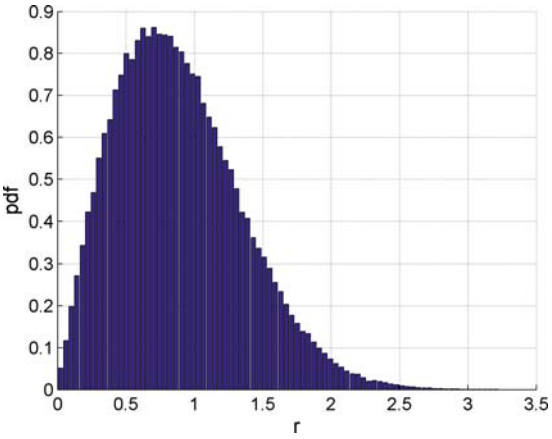


Fig. 15.12 OFDM's envelope histogram

Figure 15.12 shows the histogram of the WLAN-OFDM envelope. This shape matches pretty well a Rayleigh distribution:

$$pdf(r) = \frac{r}{\sigma^2} \exp\left(\frac{-r^2}{2\sigma^2}\right) \tag{15.5}$$

In order to generate the phase modulated carrier in Fig. 15.11, the oscillating frequency of a VCO (or DCO) is slightly changed around its nominal value. Instantaneous oscillating frequencies are calculated as the difference between consecutive phase samples normalized to the sampling period. This operation is carried out by the phase-to-frequency (PF) converter block. Instantaneous frequency samples are used to control the control voltage of a VCO, or the digital word in a DCO. The frequency modulation is performed using one or two point modulation within the PLL, with the consequent saving on hardware. The phase signal is reconstructed using a sample-and-track filter (sample-and-hold for the instantaneous frequency deviation), which is a reasonably good reconstruction. Figure 15.13 shows the histogram of the instantaneous frequency of a WLAN-OFDM signal. The picture shows that very large instantaneous frequencies are possible (usually near zero-crossings). In order to have an acceptable signal to quantization noise ratio the largest frequency values have to be clipped. This implies that the VCO/DCO will not add/subtract phase fast enough, and therefore the achieved phase modulation will deviate from the ideal one for a few samples.

The main drawback of the above described architecture is that the bandwidth for which DC-to-DC converters present good efficiencies is rather limited. Larger modulation bandwidths require higher switching frequencies for the same spur suppression. This however, increases switching losses. Figure 15.14 shows the envelope signal bandwidth for WLAN-OFDM. This bandwidth is orders of magnitude above what state-of-the-art DC-to-DC converters can offer efficiently, which makes the architecture in Fig. 15.11 unpractical for WLAN applications.

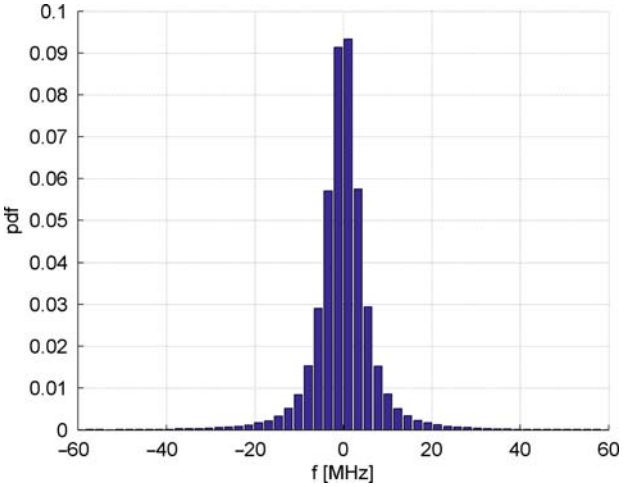


Fig. 15.13 OFDM's instantaneous frequency histogram

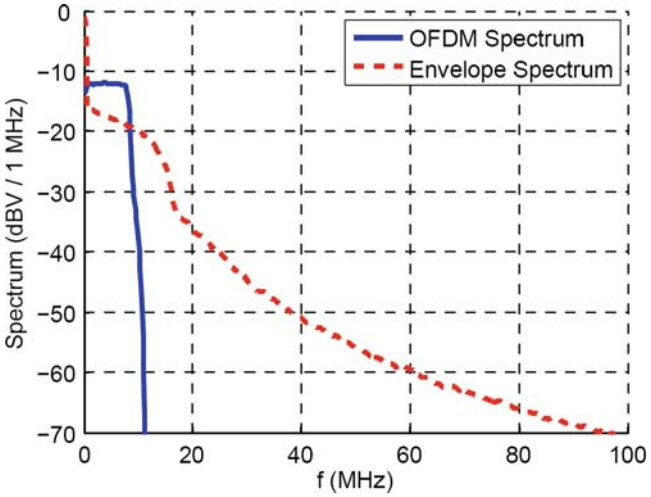


Fig. 15.14 OFDM's envelope spectrum

15.3.1 Direct Digital Polar Transmitter

Next, we present a multi-mode fully-digital polar transmitter. The block diagram of such transmitter is shown in Fig. 15.15. The phase modulated carrier is obtained just as before, but the amplitude is modulated using an envelope DAC (EnvDAC), or digital envelope modulator. Here, the envelope samples, represented as a digital bus, directly control the envelope modulation. Since there is no reconstruction filter in

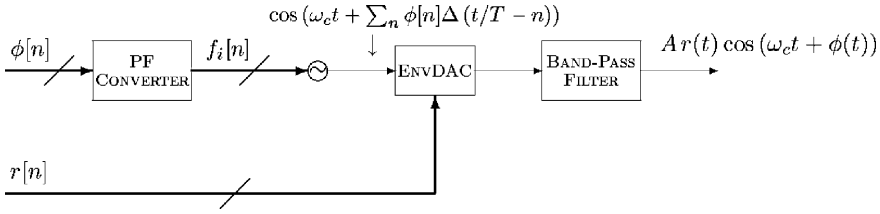


Fig. 15.15 Polar transmitter with digital frequency and envelope modulation

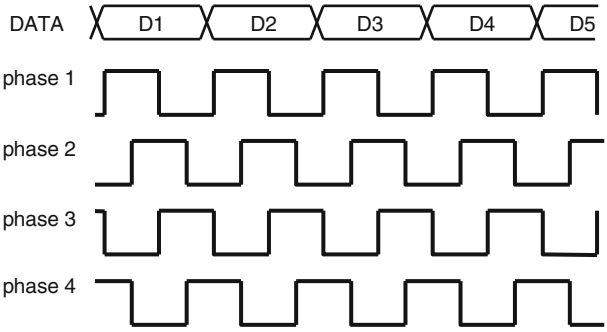


Fig. 15.16 4-phase clocking: same data is clocked by four clock edges

the envelope path, a band-pass filter might be required to attenuate the aliases at the output of the modulator. The envelope sampling frequency has to be high enough to push images outside the application band.

This architecture has virtually no bandwidth limitations and it allows for multi-standard transmission. It combines the envelope modulation function with the PA function, allowing for digital pre-distortion to compensate for saturation effects. This allows for good error vector magnitude (EVM) values without sacrificing too much in efficiency. However, this approach has the inconvenience of generating aliases at RF due to the sample-and-hold envelope reconstruction. Fortunately, the problem can be mitigated by using multi-phase clocking, which pushes strong images away without having to increase the actual data rate. This greatly relaxes the band-pass filter requirements.

15.3.2 Multi-phase Clocking

Multi-phase clocking has been reported in [8] as L-fold interpolation. Figure 15.16 illustrates the concept when using four clock-phases. The same data is fed to four identical DAC converters, but each DAC uses a clock signal with a different phase. Then the DAC outputs are added. The use of four phases emulates a 4-times higher clock without the need for accurate data interpolation. This significantly reduces

the power of the first three aliases. When using four phases, the attenuation vs. frequency of the DAC images is given by the following formula:

$$Att(f) = \sin c\,(f\,T_s)\cos\left(\frac{\pi f\,T_s}{4}\right)\cos\left(\frac{\pi f\,T_s}{2}\right) \tag{15.6}$$

15.3.3 IC Implementation

In the following, we focus on the implementation of an envelope modulator in CMOS65. The target applications are Bluetooth and IEEE 802.11 g operating at 2.4 GHz, but the same architecture can be used for IEEE 802.15.4 at 868/915 MHz and 2.45 GHz.

The basic principle behind the proposed envelope modulator is shown in Fig. 15.17. In here an array of eight binary-weighted, triple-cascade devices is shown. The output current is controlled by adding more or less stages in parallel to the output by switching *on* or *off* the middle transistor. As a consequence, the output current is proportional to the envelope code (r8 . . . r1). The lower devices receive at their input the phase modulated carrier, so the output is also proportional to the RF input. The output RF power can be controlled by both changing the digital word and the RF input power. The upper thick-oxide transistor is added for reliability, because the swing at the output can be close to two times the supply voltage during envelope peaks.

In order to guarantee a monotonic behavior of the output envelope versus input code, the actual implementation uses a thermometer coded array. This is shown in Fig. 15.18. Moreover the design is pseudo-differential to reduce ground bouncing due to the package bondwire inductance to ground.

The unit-cell consisting of a triple cascade stage and an AND-OR gate is shown on the right hand side of Fig. 15.18. As mentioned earlier, the upper cascade is a thick-oxide (50 A°) device and its main function is to protect the two lower transistors from the large voltage swing at the output (close to 5 V for high output powers).

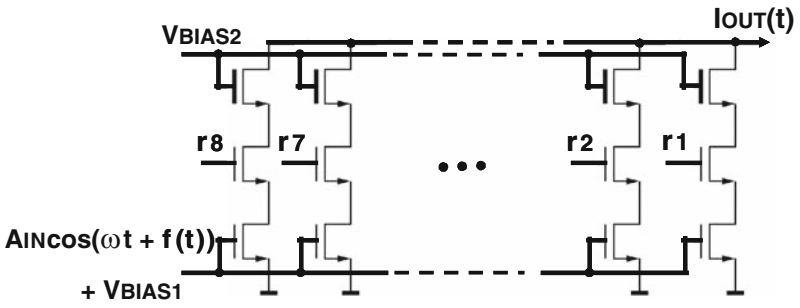


Fig. 15.17 Conceptual view of the digital envelope modulator

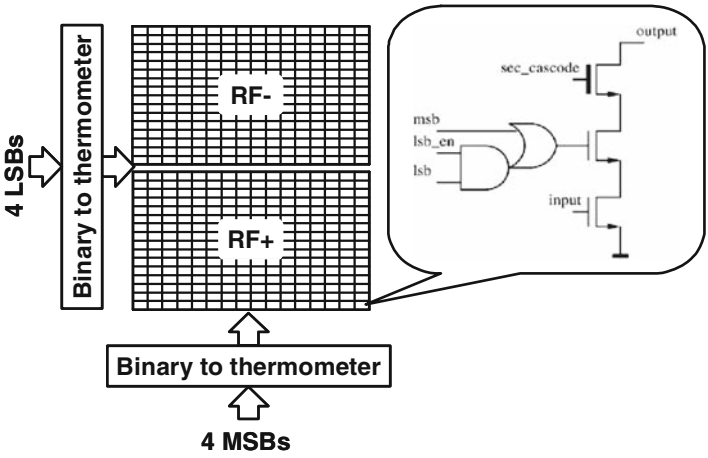


Fig. 15.18 Thermometer-coded implementation of the digital envelope modulator

The gate of the thick–oxide device is connected to a 2 V external reference. The middle transistor is used as a switch and its gate voltage is either 0 or 1.2 V depending on the values of the digital inputs *msb*, *lsb_en*, and *lsb*. The gate of the lower transistor, or input, is driven by the RF signal plus biasing voltage. Both input RF swing and biasing point can be adjusted to change output power and/or mode of operation. The AND-OR gate is part of the binary-to-thermometer decoding which translates the 8-bit envelope word to its equivalent 255-bit thermometer-encoded value. As shown in the figure, 256 unit-cells are arranged in a 16×16 matrix. The 8-bit envelope input is split into 4 MSB's and 4 LSB's. The MSB's determine how many columns are *on*, while the LSB's select how many unit-cells from the column with a valid *lsb_en* signal are *on*. Two of these matrices are used to create a pseudo-differential structure with common *msb*, *lsb_en*, and *lsb* signals, but 180°-shifted RF inputs. The layout of the unit cell is very compact. The drain and source of the input and switch transistors have been merged to reduce size and parasitics. The drain of the switch transistor and the source of the thick-oxide have been also merged, but their sizes cannot be reduced too much due to manufacturability rules concerning the distance of a thick-oxide gate to a standard thickness gate.

To lower the sampling frequency requirements a 4-phase clocking scheme has been implemented. The complete envelope modulator design consists of 4 pseudo-differential matrices along with their binary-to-thermometer decoders (as the one shown in Fig. 15.18). All matrices share the same differential RF input, envelope code and differential RF output, but their binary-to-thermometer converters use 90° degrees-shifted clock signals. This is shown on the left side of Fig. 15.19.

On the right hand side of picture Fig. 15.19 the output envelope versus sampling time is shown. The black trace represents the ideal output envelope. The dashed blue line shows the output envelope when using conventional single clock phasing. Here a large jump takes place every clock period and then the value is held till the

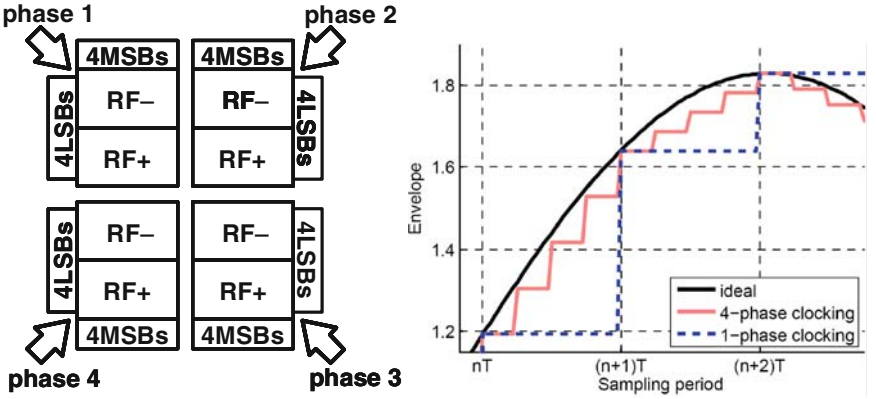


Fig. 15.19 Multi-phase clocking implementation

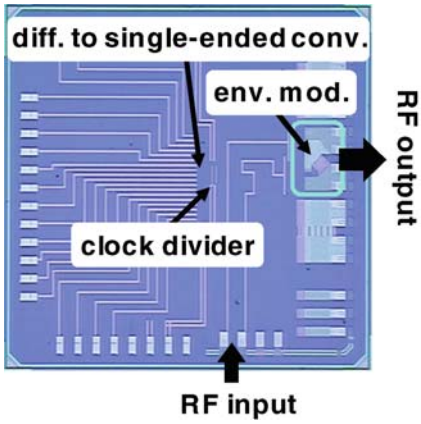


Fig. 15.20 Chip micrograph of envelope modulator

next sampling moment. The solid red line shows what happens when using four delayed phases of the same clock. At the beginning of the cycle, one of the matrices changes its output to one forth of the next envelope code. One quarter of the clock period later, the second matrix changes, so the envelope level at that time is half the previous envelope sample plus half the next envelope sample. The cycle ends when the four matrices hold the same data value for one quarter of the clock period. In this way, linearly interpolated values in between two envelope samples are obtained. The interpolated values are not ideal, but the approximation is good enough to substantially reduce the power of the near-by images without having to spend power calculating the exact values.

The total chip micrograph is shown in Fig. 15.20. The eight input envelope bits are differential. The input differential clock runs at two times the data rate and a clock divider is used to produce the four phases used to clock the data. The modulator area itself is smaller than 0.1 mm^2. In the modulator, a metal-5 ground plane is used to isolate the RF output from the RF input. This minimizes unwanted carrier

leakage due to crosstalk. The chip is packaged in a 48-pin HVQFN package. Each RF output is connected to two bondwires to reduce the series inductance, while six downbonds are used to have good grounding.

15.3.4 Measurement Results

In the following we show the measured static behavior of the envelope modulator at 2 GHz. In a first measurement the input envelope code is swept from 0 to 255 and for every code the output carrier amplitude is measured. This sweep can be performed for different input RF power values, since the output power also depends on the input RF power and biasing conditions. The output RF-amplitude vs. envelope code for two different input powers and biasing conditions is shown in Fig. 15.21 at the left. The solid blue curve shows the result for a maximum 255-code continuous wave (CW) output power of 9.2 dBm. In this case the performance is reasonably close to the ideal behavior, which is a straight line. The 1-dB compression code is 251. The red dashed curve shows the measurements for a maximum CW output power of 24.5 dBm. In this case, the 1-dB compression code decreases to 105.

At the right in Fig. 15.21 the phase-shift added by the envelope modulator in function of the code is shown. For the low-power setting, the characteristic remains close to the ideal behavior, a constant value, while for the large output power the phase-shift becomes much more significant, with a maximum variation of about 9 degrees.

Since the envelope modulator has a digital input, the code-to-AM can be digitally pre-distorted to achieve a more ideal behavior. The code-to-PM characteristic can also be digitally pre-distorted in the phase modulator path.

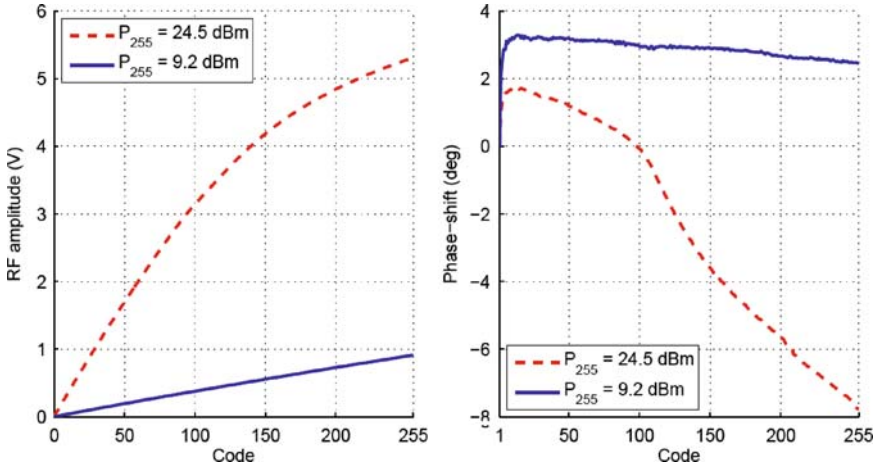


Fig. 15.21 Left: static code-to-AM measurement and right: static code-to-PM measurement

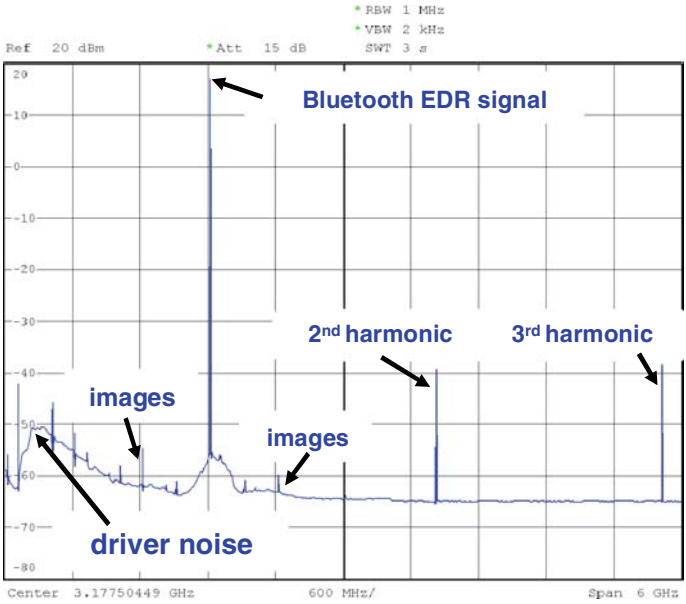


Fig. 15.22 Bluetooth EDR measurement

Figure 15.22 presents the measured results when using the envelope modulator to generate a Bluetooth Enhanced Data Rate (EDR) signal. The bandwidth of the signal is 1 MHz, with a data rate of 3 Mb/s achieved using eight differential phase shift keying. The peak to average ratio of this modulation is about 3.3 dB. For this measurement an envelope sampling frequency of 300 MHz was used. The high oversampling ratio combined with multi-phase clocking means that the images are heavily attenuated. The measured average output power is 19.7 dBm, for a mean EVM of 5%. Note that this is much lower than the standard requirement of 13%. The drain efficiency is 26%. The noise coming up at low frequencies is generated by the external driver used to amplify the input phase-modulated carrier.

Next, the envelope modulator is used to generate an OFDM-WLAN signal. The envelope sampling frequency is 200 MHz. Figure 15.23 shows the partially cancelled first, second and third aliases, and the strong fourth alias. This measurement matches with the predicted attenuation when using 4-phase clocking. The spectrum is asymmetrical because the fourth alias on the right is attenuated by the filtering introduced by the matching network. The spectrum complies with the IEEE spectrum mask with 8 dB margin.

The measured burst power is 16.7 dBm, which is enough for many existing applications. This power is obtained at a drain efficiency of 24%. This value is comparable to the efficiency achieved by external PA's, designed in more favorable technologies. The measured EVM is a competitive 2.7%, which is much better than 5.6% required by the Standard. The measured IQ offset represents the measurement equipment limit, which turns out to be −55 dB.

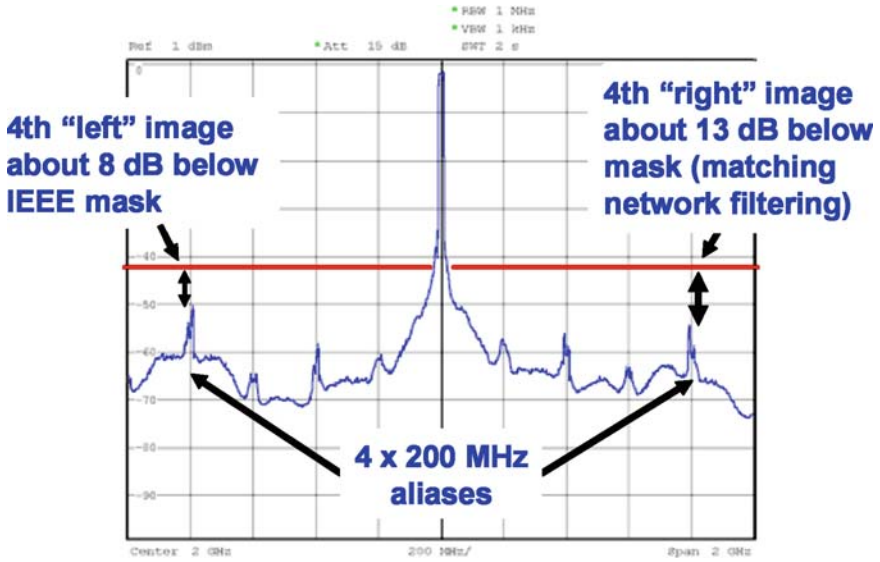

4th "left" image about 8 dB below IEEE mask
4th "right" image about 13 dB below mask (matching network filtering)
4 x 200 MHz aliases

Fig. 15.23 WLAN-OFDM spectrum measurement

15.4 Conclusions

In this paper we first describe a new passive voltage mixer driven by 25%-duty-cycle LO signals. Based on this voltage mixer concept, we demonstrate a complete low-power low-EVM and SAW-less WCDMA modulator in a 45 nm CMOS process. This modulator can also be applied to other 2.5, 3 and 4G applications.

Secondly the implementation of a direct digital polar transmitter for WLAN and BT is described. It combines digital-to-analogue conversion, up-conversion and power amplification, allowing for a fully-integrated transmitter solution. It is shown how the use of multi-phase clocking reduces aliases to acceptable levels without the need for accurate envelope-data interpolation. The implementation delivers sufficient output power for most WLAN and BT implementations with a comfortable EVM, transmit spurious emission margin and competitive efficiency.

References

1. Staszewski, R.B. et al., "All-Digital TX Frequency Synthesizer and Discrete-Time Receiver for Bluetooth Radio in 130-nm CMOS", IEEE J. Solid-State Circuits, Vol. 39, No. 12, December 2004.
2. A. Shameli, A. Safarian, A. Rofougaran, M. Rofougaran, and F. de Flaviis, "A Novel DAC Based Switching Power Amplifier for Polar Transmitter", IEEE Custom Integrated Circuits Conference (CICC), pp. 137–140, 10–13 September 2006.

3. D. Papadopoulos and Q. Huang, “A Linear Uplink WCDMA Modulator with −156 dBc/Hz Downlink SNR,” ISSCC Dig. Tech. Papers, pp. 338–339, February 2007.
4. C. Jones, B. Tenbroek, P. Fowers, et al., “Direct-Conversion WCDMA Transmitter with −163 dBc/Hz noise at 190 MHz Offset,” ISSCC Dig. Tech. Papers, pp. 336–337, February 2007.
5. A. Mirzaei, H. Darabi, “A Low-Power WCDMA Transmitter with an Integrated Notch Filter,” ISSCC Dig. Tech. Papers, pp. 212–213, February 2008.
6. X. He, J. van Sinderen, “A 45 nm Low-power SAW-less WCDMA Transmit Modulator Using Direct Quadrature Voltage Modulation”, ISSCC Dig. Tech. Papers, pp. 120–121, Feb. 2009.
7. C. Mayer, et al., “A Robust GSM/EDGE Transmitter Using Polar Modulation Techniques”, The European Conference on Wireless Technology 2005, pp. 93–96.
8. Y. Zhou, J. Yuan “A 10-bit Wide-band CMOS Direct Digital RF Amplitude Modulator”, IEEE J. Solid-State Circuits, vol. 38, issue 7, pp. 1182–1188, July 2003.

Chapter 16
Challenges for Mobile Terminal CMOS Power Amplifiers

Patrick Reynaert

Abstract Several PA efficiency enhancement techniques, tailored towards CMOS implementation, are discussed. It will be shown how the combination of reconfigurable PA circuitry, together with the fast processing power of CMOS, can result in novel TX and PA architectures that achieve a higher efficiency when amplifying amplitude modulated signals

16.1 Introduction

Although CMOS, allowing an integration of all baseband and RF functionality on a single chip, is the preferred technology to implement a transceiver for wireless consumer applications, the integration of the RF-PA in CMOS is non-obvious. Recently, the integration of Watt-level constant-envelope PAs became possible thanks to the introduction of novel power combining architectures [1] based on transformers. Today, research is shifting its focus on linear PAs. The move towards wireless standards such as WCDMA, WLAN, WiMAX, ... has created a strong demand for better PA efficiency when amplifying and transmitting amplitude modulated signals.

This paper will emphasize on the use of digital signal processing, together with flexible/reconfigurable CMOS PA architectures, to achieve linear and efficient RF amplification in CMOS. Since CMOS is ideally suited for digital signal processing and control, it can easily handle and steer a complex PA architecture with many *knobs and handles*. It is exactly in this field that the integration of an RF PA in CMOS makes sense. This paper will emphasize some of the recent research results that have clearly demonstrated this successful marriage between a reconfigurable PA architecture and digital conversion techniques.

P. Reynaert (✉)
Ercan Kaymaksüt and Brecht François, K.U.Leuven ESAT-MICAS, Kasteelpark Arenberg 10, B-3001 Leuven Belgium
e-mail: patrick.reynaert@esat.kuleuven.be

A.H.M. van Roermund et al. (eds.), *Analog Circuit Design: Smart Data Converters, Filters on Chip, Multimode Transmitters*, DOI 10.1007/978-90-481-3083-2_16,
© Springer Science+Business Media B.V. 2010

16.2 Efficiency Improvement Techniques

Let us first go back to the basic problem. In a Class A amplifier, the DC current through the transistor is fixed and set by the gate bias voltage. The load resistor is chosen so that, for a given maximum current swing, the voltage swing at the transistor drain terminal is maximized. This indeed results in the maximum output power and maximum efficiency of 50%.

When that same amplifier transmits a lower output power, the efficiency of the PA will degrade linearly with the output power. Indeed, the DC current consumption and supply voltage do not change when amplifying AM modulated signals. As such, the DC power consumption remains constant and the efficiency is therefore proportional to the output power:

$$\eta \propto P_{OUT} \tag{16.1}$$

meaning that at 6 dB power back-off (a typical number for modern communication systems), the efficiency of the Class A PA is reduced by a factor of four. This is a major drawback of the Class A amplifier. The efficiency-versus-output power curve is far from optimal for the amplification of amplitude modulated signals, as present in modern communication systems like WLAN, WiMAX, WCDMA and so on.

The cause of the efficiency reduction at lower output power is clear: the constant power dissipation in the transistor. The solution is also trivial at first sight: when the output power is reduced, the supply voltage and/or the transistor bias current have to be reduced as well. An alternative might be to increase the load resistor when less output power is required. This forms the basis for many efficiency improvement techniques (Doherty, adaptive biasing, envelope tracking . . .).

There are various ways to implement the above ideas and this has lead to the wide variety of efficiency improvement and linearization techniques that exist for power amplifiers. In what follows, we will discuss some of the implementations issues, especially focused on CMOS.

16.3 Changing the RF Path

From the previous section, it became clear that we need to *change something* in the PA bias settings or the load resistance (as seen by the PA) to improve the efficiency at power back-off levels.

A well-known PA already achieves this goal: the Class B PA [2]! First of all, the peak efficiency of the Class B PA is higher compared to Class A. This has to do with the reduced voltage-current overlap when looking at the transistor drain-current and drain-voltage (for a MOS). But more important, the Class B PA also achieves a more favorable efficiency-versus-output power curve which is proportional to the square root of the output power, instead of decreasing linearly with decreasing output power as in Class A.

In an ideal Class B amplifier, the quiescent or bias transistor current is zero. The DC drain current, and hence the DC current consumption, is proportional to the input signal amplitude, as shown in Fig. 16.1.

As such, the DC current consumption scales linearly with the input signal. Assuming that the supply voltage is kept constant, also the DC power consumption of the PA scales linearly with the input signal. The output power obviously scales quadratic with the input signal amplitude and as such, the efficiency is proportional to the square root of the output power.

$$\eta \propto \sqrt{P_{OUT}}$$

In other words, at the typical 6 dB power back-off point, the efficiency is *only* reduced by two in a Class B PA which is more favorable than Class A (see Fig. 16.2).

Although beneficial for the efficiency, the change of the DC transistor current with the input amplitude will cause severe non-linearities (varying output conductance, capacitance and transconductance). There exists other ways to implement the Class B principle more efficiently, and this will be discussed in the next section.

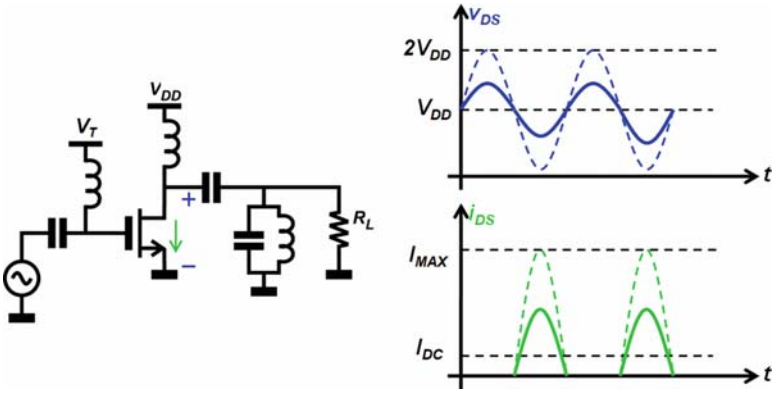


Fig. 16.1 Class B schematic and waveforms

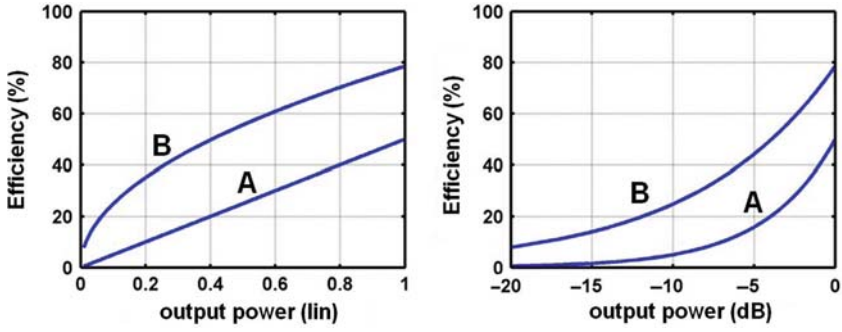


Fig. 16.2 Efficiency versus output power

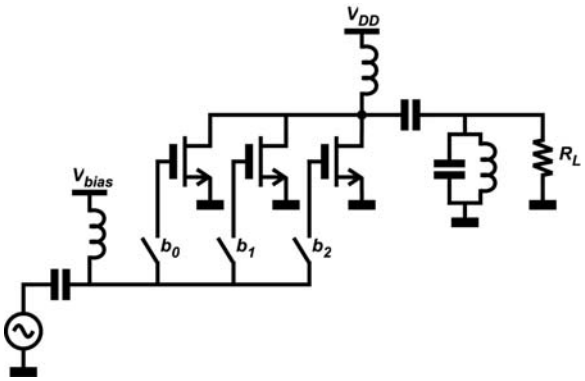


Fig. 16.3 Discrete type Class B PA

16.3.1 Discrete Class B

Instead of scaling the amplitude of the input RF signal to change the RF output current, one can also change the transconductance of the PA by changing the total transistor gate-width. This, of course, cannot be done in a continuous manner. Instead, we need to partition the *big* nMOS transistor of Fig. 16.1 in a parallel structure of *many* small transistors which can individually be turned on or turned off. The concept of this approach is shown in Fig. 16.3. The amount of RF current delivered to the output is now determined by the digital word $[b_0 b_1 b_2]$. This can be binary or thermometer code or a combination of both for the highest dynamic range and linearity.

For the efficiency, the important thing is that the DC current consumption scales with the amount of current that flows towards the output, similar to Class B operation. And as such, the efficiency of such a topology will be proportional with the square root of output power, like in a normal Class B PA.

An example of such an approach was presented in [3,4] and more recently in [5].

16.3.2 Power Combining

As it was indicated before, the efficiency of a Class A PA degrades linearly with output power due to the constant current consumption, power supply and load impedance. Scaling the DC current consumption with respect to the output current (Class B) is a first step to improve the efficiency at power back-off. This renders a square-root behavior. One could further improve the efficiency at power back-off by increasing the load resistance when less output power is required [2, 6]. Obviously, this requires a *reconfigurable* or *electronic steerable* impedance matching network.

This demand puts a strong requirement on the passive elements as one would need a high-quality varactor. An alternative is to make use of power combining networks.

When using a non-isolated power combiner, such as a transformer-based combiner [7], the different input ports, and hence the PAs themselves, are coupled to each other. As such, one can electronically change the impedance, seen by each PA, by turning-on or -off one or more PAs. This is very similar to *active load pull*. It is important that a non-isolated combiner is used as opposed to e.g. a Wilkinson power combiner, as otherwise the input impedance will not change which is crucial to improve the efficiency at power back-off.

Assume we have a system of two power amplifiers with two 1:1 transformers as shown in Fig. 16.4.

When the two amplifiers are simultaneously active, the transformed load impedance R_{IN}, seen by each PA, is equal to $R_L/2$. When one amplifier is turned-off, and assuming the secondary winding of the inactive section can somehow be short-circuited, the load impedance increases to R_L, which is two times as large as before. Of course, there are some implementation issues involved with *turning-off* a PA and *short-circuiting* the transformer winding, but that is not the focus of this paper. Since the load impedance R_{IN} has doubled, the PA input voltage needs to be reduced by a factor two to have the same maximum voltage swing at the drain. Furthermore, the output power is reduced by 6 dB since only one PA delivers power to the load (-3 dB) and the power that the active PA delivers is reduced by 2 (-3 dB) because of the higher load impedance. All this means that at 6 dB power back-off, the efficiency is the same as the peak efficiency because of two reasons. First of all, the PAs operate in Class B, meaning that their DC current consumption scales with the input power. Secondly, the increase of the load resistor at power back-off allows a full voltage swing at the drain, resulting in the same PA efficiency as at full output power. This results in a set of efficiency-versus-output power curves as shown in Fig. 16.5. An example of this principle can be found in [8].

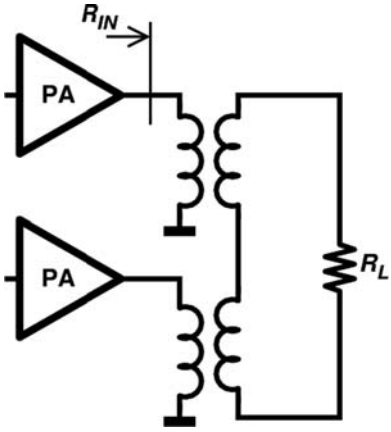


Fig. 16.4 Power combining topology with two power amplifiers

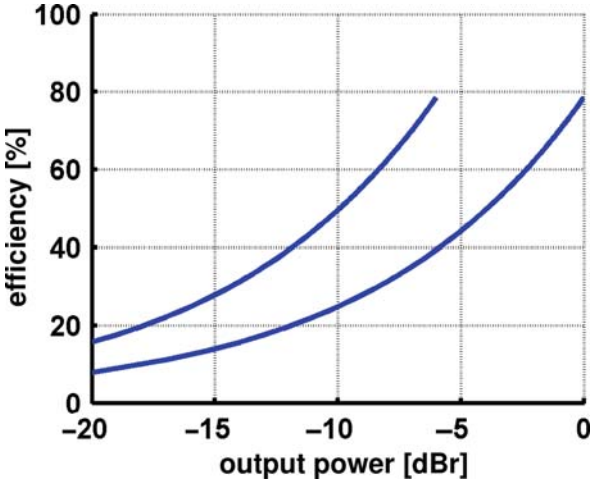


Fig. 16.5 Efficiency versus output power for a 2-way combining Class B PA with transformers

16.4 Changing the BB Path

So far, we have only discussed how to change the PA properties through the RF signal path. Other approaches exist that are based on the baseband path. Indeed, one could change the supply voltage or the gate bias voltage of the transistor when the RF envelope is lower. In CMOS, such a control scheme can be made independent from the RF path since the envelope signal can be obtained directly from the I(t) and Q(t) signals in the transmitter, rather than extracting the envelope from the modulated RF signal.

Envelope tracking, where the supply voltage of a Class AB PA is following or tracking the envelope signal, has successfully been applied to improve the efficiency at power back-off levels. Envelope tracking, combined with Class B operation, has the potential to achieve a flat efficiency-versus-output power curve, at least in theory.

Envelope tracking is an efficiency improvement technique where the RF PA still needs to have sufficient linearity to meet the spectral mask and EVM specifications. On the other hand, one could also go for a switching-type RF PA where the changes of the supply voltage are essentially modulating the output RF carrier. This is referred to as polar modulation or supply modulation. The RF PA now works as a switching type PA where the amplitude of the output RF signal is proportional to the supply voltage. The switching PA then behaves as a mixer. In such a linearization scheme, the linearity and efficiency requirements are completely shifted to a low-frequency block: the supply modulator (see Fig. 16.6).

The amplitude or envelope modulator, indicated by LF-PA in Fig. 16.6, can be implemented as an LDO regulator, as shown in Fig. 16.7. [9]. In other words, a series-pass transistor is used to determine the supply voltage delivered to the RF-PA. The advantage of this approach is the simplicity and wideband behavior of this approach.

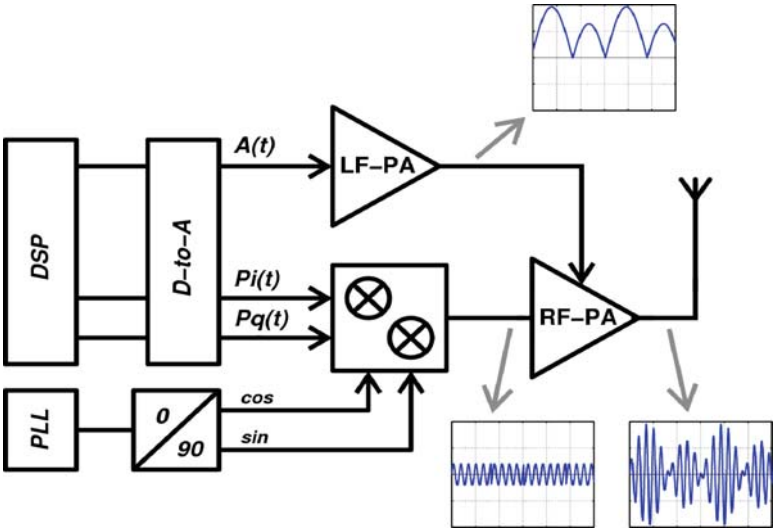


Fig. 16.6 Simplified polar transmitter

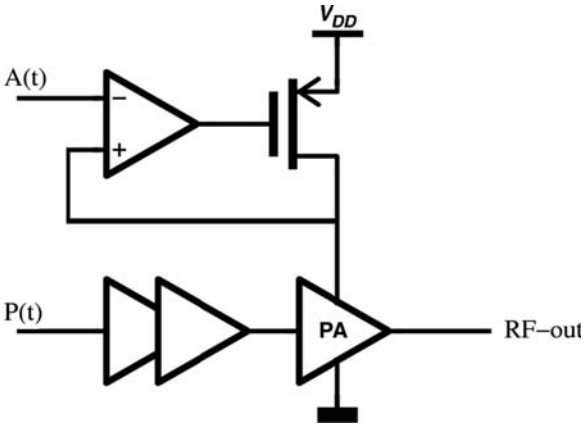


Fig. 16.7 LDO-type of amplitude modulator

A drawback is the power dissipation across the LDO pMOS transistor which reduces the efficiency of the entire solution at power back-off. Since the switching RF-PA maintains its high efficiency, even when the power supply is reduced, the efficiency becomes -again- proportional to the square root of the output power. In other words, this topology behaves as a Class B PA.

There is an important difference however with a *normal* RF Class B PA. Since all linearity requirements are shifted to baseband, the linearity of this topology is superior compared to that of a *normal* RF Class B power amplifier. As such, this solution will achieve higher peak and average efficiency compared to a Class B PA since the PA can be operated closer to the saturated output power [10].

One could also implement the amplitude modulator as a switching-type modulator like a buck or boost DC/DC converter. Such converters need an LC filter to reconstruct the low-frequency signal from the switching square-wave waveform. Given the low-frequency nature of these LC filters requires a rather large or *bulky* inductor. This inductor needs to have low-loss, high current capability and high self-resonance frequency. Obviously this inductor becomes expensive.

16.4.1 Digital Polar

We have already noted that the switching type amplifier behaves as a mixer. Imagine that we could *push* the low-pass LC filter through the RF PA. This would mean that the low-pass filter gets up-converted by the switching nature of the PA, which behaves as a mixer. As a consequence, the low-pass filter becomes a band-pass [11]. This is very similar to the synthesis of band-pass filters in classical filter theory designs. The architecture now looks as in Fig. 16.8 with a band-pass filter between the PA and the load. The low-frequency switching modulator turns on and off the switching PA. The result is a series of RF bursts, generated by the multiplication of a $\Delta\Sigma$, PWM or other kind of modulator signal with a phase-modulated RF signal.

Obviously, turning on and off a PA by turning on/off the supply voltage is not an efficient approach. The architecture can indeed be further improved by turning on/off the gate signal, as shown in Fig. 16.9.

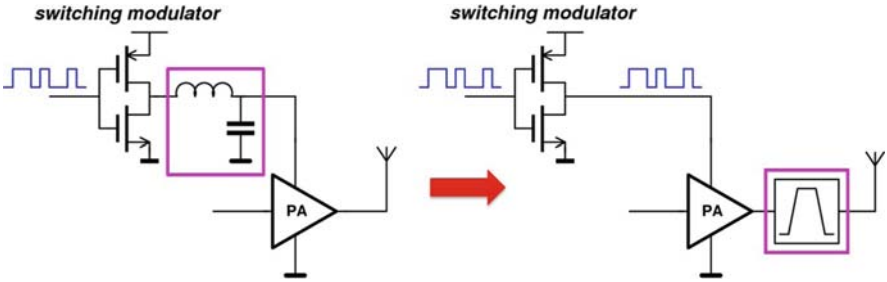


Fig. 16.8 Pushing the low-pass filter through the PA makes it a bandpass filter

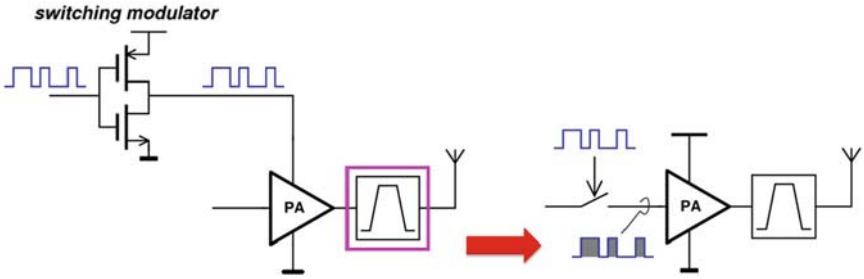


Fig. 16.9 Digital polar modulation

The band-pass filter at the PA output deserves a little more attention. Indeed, just as the inductor in the low-pass filter had some stringent requirements, the same is true for the band-pass filter at the PA output. Often, such a filter is already there to lower the noise of the transmitter in the receive band, although recent trends have indicated the possibility to make this filter superfluous. Furthermore, this filter still needs to have low loss as the loss of this filter is directly reflected in the PA efficiency.

Another important issue is the out-of-band energy. The low-frequency modulator generates spectral components at multiples of the modulator sample frequency. This out-of-band energy, mainly concentrated at the sample frequency, might violate the spectral mask requirements of the specific wireless standard. Furthermore, if this out-of-band energy is dissipated, the efficiency of the PA degrades to that of a Class B amplifier [12, 13]. These additional requirements result in the need for a high out-of-band filter impedance, low-loss (efficiency requirements) and steep roll-off (spectral mask requirements). Obviously, this will increase the cost of the filter. In CMOS, this topology still holds promise if the burst sample frequency can be kept high enough which relaxes the filter requirements. Furthermore, coding can be applied that minimizes the out-of-band noise. An example of a system based on this approach was presented in [12, 14].

16.5 Conclusions

Integrating a Watt-level constant envelope PA is CMOS has become feasible. The next step will be to improve the efficiency of a linear PA topology. Here, one has to make use of the high complexity that can be mastered with CMOS. Signal processing and control is the natural habitat of CMOS. Reconfigurable PA circuits steered and controlled by fast CMOS logic, will render new TX and PA architectures that achieve higher efficiency at power back-off.

References

1. I. Aoki, S. Kee, D. Rutledge, and A. Hajimiri, “Fully-Integrated CMOS Power Amplifier Design Using the Distributed Active Transformer Architecture,” IEEE Journal of Solid-State Circuits, vol. 37, no. 3, pp. 371–383, March 2002.
2. S. C. Cripps, “RF Power Amplifiers for Wireless Communications”, Artech House, 2006.
3. A. Kavousian, D. K. Su, M. Hekmat, A. Shirvani and B. A. Wooley, “A Digitally Modulated Polar CMOS Power Amplifier With a 20-MHz Channel Bandwidth”, IEEE JSSC, vol. 43, no. 10, pp. 2251–2258, October 2008.
4. C. D. Presti, F. Carrara and A. Scuderi, “A High-Resolution 24-dBm Digitally Controlled CMOS PA for Multi-Standard RF Polar Transmitters”, ESSCIRC 2008.
5. S. Kousai and A. Hajimiri, “An Octave-Range Watt-Level Fully Integrated CMOS Switching Power Mixer Array for Linearization and Back-Off Efficiency Improvement, ISSCC 2009 Digest of Technical Papers, pp. 376–377.

6. F. H. Raab, “High-Efficiency Linear Amplification by Dynamic Load Modulation”, IEEE IMS 2003, pp. 1717–1720.
7. D. Chowdhury, P. Reynaert and A. M. Niknejad, “Transformer Coupled Power Amplifier Stability and Power Back-off Analysis”, in IEEE Transactions on Circuits and Systems – II, vol. 55, pp. 507–511, June 2008.
8. G. Liu, T.-J. K. Liu and A. M. Niknejad, “A 1.2 V, 2.4 GHZ Fully Integrated Linear CMOS Power Amplifier with Efficiency Enhancement, Proceedings of IEEE 2006 Custom Integrated Circuits Conference, pp. 141–144.
9. P. Reynaert and M.S.J. Steyaert “A 1.75 GHz GSM/EDGE Polar Modulated CMOS RF Power Amplifier”, Proceedings of IEEE International Solid-State Circuits Conference 2005, pp. 312–313.
10. P. Reynaert and M. Steyaert, “A 1.75-GHz Polar Modulated CMOS RF Power Amplifier for GSM-EDGE,” in IEEE Journal of Solid-State Circuits, vol. 40, pp. 2598–2608, December 2005.
11. P. Reynaert and M. Steyaert, “RF Power Amplifiers for Mobile Communications,” Springer, the Netherlands, 2006.
12. W. Laflere, M.S.J. Steyaert and J. Craninckx, “A Polar Modulator Using Self-Oscillating Amplifiers and an Injection-Locked Upconversion Mixer,” IEEE Journal of Solid-State Circuits, vol.43, no.2, pp.460–467, Feb. 2008.
13. P. Reynaert, W. Laflere, M. Steyaert and J. Craninckx, “Self-oscillating RF Amplifiers”, GigaHertz Symposium 2008, Goeteborg, Sweden, March 2008.
14. J. Stauth and S. R. Sanders, “A 2.4 GHz, 20 dBm Class-D PA with Single-bit Digital Polar Modulation in 90 nm CMOS”, in IEEE CICC 2008, pp. 737–740.

Chapter 17
Multimode Transmitters with ΔΣ-Based All-Digital RF Signal Generation

A. Frappé, A. Kaiser, A. Flament, and B. Stefanelli

Abstract This paper presents an all-digital approach to the generation of the modulated radio-frequency carrier and its application to a multimode transmitter in today's communication systems and draws a possible picture of tomorrow's systems. We will first analyze how digital transmitters will progressively replace their analog counterparts and what are the main issues associated with this trend. The combination of ΔΣ modulation and digital mixing is proposed as an innovative approach enabling multimode operation of transmitters with low power consumption and chip area, easy configurability, and good performance. A 90 nm CMOS chip has been designed to demonstrate the feasibility of the concept and its potential in multimode transmitter architectures. Techniques such as redundant arithmetic and non-exact quantization are used in the high-speed ΔΣ modulator implementation. Furthermore, approaches to antenna filtering using BAW filters and reconfigurable semi-digital RF FIR filters will be introduced. Finally, a review of recent outstanding transmitter designs will allow a comparison between the presented approach and architectures based on digital-to-RF conversion.

17.1 Introduction

Everything is now wireless or tends to become wireless. Needs from consumers are satisfied by the evolution of smart phones, personal computers and other portable devices to an anywhere/anytime connectivity. As these needs range from simply talking on the phone to exchanging high-definition video streams, a large number of communication standards have been developed in time, each one with its own capabilities (GSM, EDGE, GPRS, WCDMA, HSDPA, WIFI, WIMAX, 802.11, 802.16 ...). Moreover, the allocated frequency bands generally differ with geo-

A. Frappé (✉), A. Kaiser, A. Flament, and B. Stefanelli
IEMN/ISEN, Lille, France

A. Frappé
Berkeley Wireless Research Center, UC Berkeley, CA, USA
e-mail: antoine.frappe@gmail.com

A.H.M. van Roermund et al. (eds.), *Analog Circuit Design: Smart Data Converters, Filters on Chip, Multimode Transmitters*, DOI 10.1007/978-90-481-3083-2_17,
© Springer Science+Business Media B.V. 2010

graphical areas. For example, the main European 3G UMTS network has allocated other frequency bands than its equivalent American standard. Recently introduced opportunistic radios could also take advantage of the available space by sensing the spectrum and automatically adapt themselves to transmit on the best available band.

Multi-standard terminals are usually made of several silicon chips, each one devoted to a particular standard and optimized for a defined bandwidth, frequency band and dynamic range. It becomes very costly and less flexible for integration, as terminals need to offer the maximum possible connectivity.

Moreover, analog RF front-ends are particularly sensitive to RF impairments. Figure 17.1 describes a traditional homodyne analog transmitter with baseband or IF digital-to-analog conversion. Typical RF impairments in the DACs and low-pass filters (LPF) include pass-band response distortions, imperfect image rejection of spectral images and I/Q imbalance. The upconversion mixer also suffers from nonlinear distortion, carrier leakage and DC offsets. All these imperfections must be corrected by either analog or digitally-assisted calibration processes.

Moving from analog dedicated front-ends to flexible digital front-ends then appears obvious in the perspective of software-defined radios. Multi-mode terminals would integrate single-chip configurable digital RF modules in a short-term future, to provide a valuable reduction in cost, but also increase flexibility, integration and reliability (aging) and remove tuning circuitries [1, 2]. Power consumption would also be reduced as the hardware can be tailored by configuration for any given standard. An example of such a digital architecture is presented in Fig. 17.2.

Issues to deal with are numerous for achieving an all-integrated digital RF transmitter. First, digital RF processing must operate at very high sample rates. Means for offering the best performance without largely increasing the power consumption are needed. Then, delivering high levels of transmit power with low supply voltage CMOS technologies is also a challenge. Finally, filtering remains an issue as

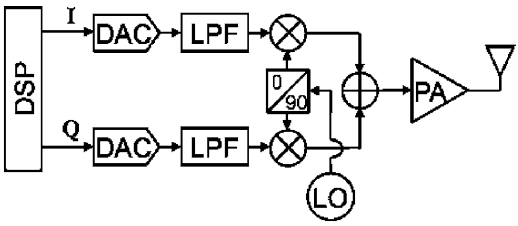


Fig. 17.1 Traditional homodyne analog transmitter

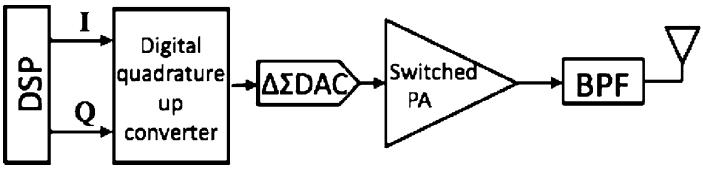


Fig. 17.2 Digital RF architecture using ΔΣ modulation and switched PA

frequency agility of antenna filters is not demonstrated yet and digital-RF architectures produce large quantization noise that needs to be removed in critical bands.

ΔΣ modulation has shown to be a good candidate for enabling digital RF transmitters and will be discussed in Section 17.2 from an architecture point of view. Section 17.3 will deal with switched power amplification by comparing voltage and current-mode approaches and describing a way to deliver high power with a low voltage supply using power combining. Section 17.4 will cover the implementation issues related to digital and mixed-signal blocks in this type of architectures, such as baseband processing, oversampling stages, ΔΣ modulators and digital RF mixers. Filtering issues will finally be reviewed in Section 17.5. In this section, recent developments on BAW filtering and RF semi-digital FIR filtering will be discussed as they could lead to valuable achievements for digital RF integrated transmitters.

17.2 ΔΣ Modulation for All-Digital RF Signal Generation

17.2.1 What Can ΔΣ Modulation Bring in Integrated Transmitters?

Oversampled ΔΣ modulation has been largely used in ADCs and DACs, mainly for audio applications, where a high dynamic range is required over a small bandwidth. Let's briefly recall the principles of oversampled ΔΣ modulation to understand how this can be applied to RF transmitters. A signal quantized on n bits and sampled at Nyquist frequency $2f_0$ has a peak Signal-to-Noise Ratio (SNR) of $6.02 \times n + 1.76$, expressed in dB, assuming the noise is uniformly distributed and uncorrelated with the signal (Fig. 17.3a). By oversampling the signal by an *OSR* factor equal to $f_s/2f_0$, the quantization noise is spread over $f_s/2$, thus decreasing the noise inside

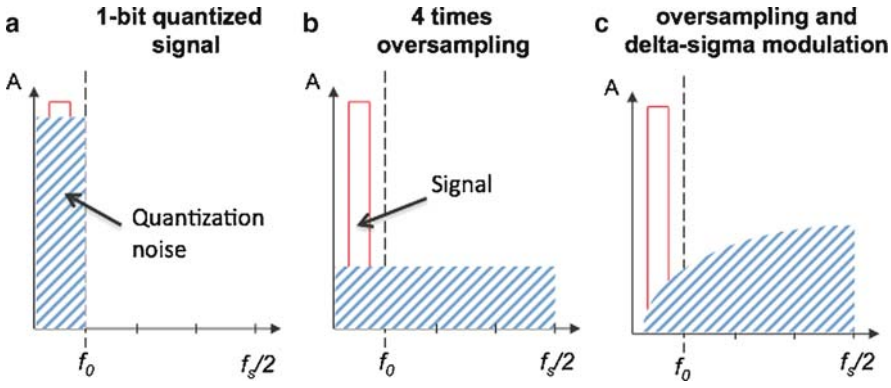


Fig. 17.3 Qualitative plots showing oversampling and delta-sigma modulation principles, **(a)** 1-b quantized signal; **(b)** four times oversampling; **(c)** oversampling and delta-sigma modulation

the band of interest (Fig. 17.3b). The SNR is improved by 3 dB with every doubling of the sampling frequency, what can be interpreted as a half bit improvement in resolution. Furthermore, the ΔΣ modulation introduces a shaping function for the noise that allows a significant noise reduction in the signal band and corresponding SNR improvement [3] (Fig. 17.3c). Note that the quantization noise has the same area in the three plots. In conclusion, a low-resolution high-speed signal is able to contain all the signal information in a defined bandwidth with a good SNR.

As advances and maturity of deep submicron CMOS technologies enable the DAC to reach higher sample rates, a DSP based on ΔΣ modulation can be envisioned to produce signals at RF frequencies. Baseband or IF DACs and conventional mixers would be progressively replaced by digital upconversion to RF and ΔΣ modulation. It can be of particular interest when using switching-mode amplifiers that can have a very high efficiency. Moreover, using a 1-bit digital output can greatly reduce the inaccuracies of high-frequency high-resolution DACs.

17.2.2 IF ΔΣ DAC

In an intermediate step to enable a digital-RF transmitter, a ΔΣ modulator and a 1-bit DAC can be used to replace the analog IF mixer stage in the transmit path, as shown in Fig. 17.4. Compared to conventional heterodyne architecture, this architecture benefits from better silicon integration, ideal IQ matching and thus a lower error vector magnitude. An implementation of this approach has been introduced in [4]. It is based on a multiplier-free digital quadrature modulation associated with a 1-bit ΔΣ modulator and a current-mode DAC. The designed 0.13 μm CMOS chip can operate at a 700 MHz clock frequency to address an IF frequency of 175 MHz. It consumes 139 mW at 1.5 V and occupies 5.2 mm^2.

17.2.3 1-bit RF DAC

The digital-to-analog limit has to be pushed towards the antenna. Our team has recently demonstrated the first 1-bit ΔΣ RF signal generator in 90 nm CMOS suitable for integration into a complete transmitter chain [5]. A 50 MHz bandwidth centered

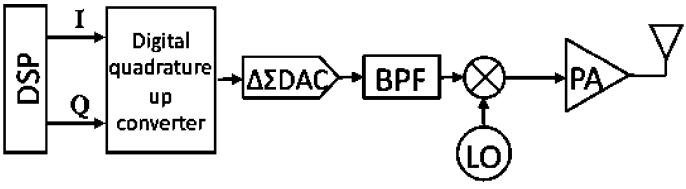


Fig. 17.4 Digital-IF architecture

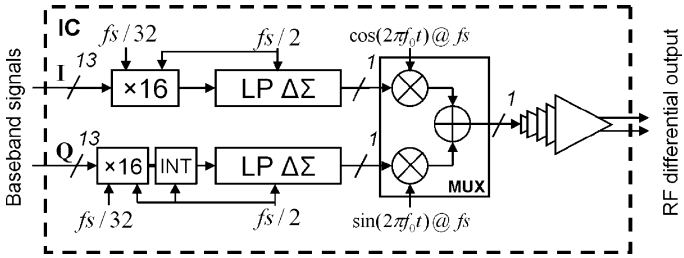


Fig. 17.5 System view of the 1-b ΔΣ RF signal generator (from [5])

on 1 GHz can be achieved when the circuit is clocked at 4 GHz. Signals up to 3 GHz can be synthesized when using the first image band. The peak output power into a 100 Ω diff. load is 3.1 dBm with 53.6 dB SNDR. The digital core consumes 49 mW at maximum clock frequency. Active area is 0.15 mm^2. A global system view of the signal generator chip is shown on Fig. 17.5.

After off-chip baseband signal processing and low-IF upconversion, the input I and Q channels are first oversampled 16 times up to $f_s/2$. The ΔΣ modulators quantize the signals into 1-bit signals and shape the quantization with a third-order transfer function including optimized zeros placement. They use a novel redundant high-speed architecture taking advantage of a 3-phase clock generated by a Delay-Locked Loop (DLL). The 1-bit digital image-reject mixer replaces the conventional mixer and upconverts I&Q signals to the desired RF carrier frequency, before being pre-amplified by sized inverters. Note that a linear interpolation stage is inserted on the Q channel to reject the image channel.

17.2.4 Multi-bit RF DAC

Recent works on multi-bit ΔΣ RF signal generation use the Digital-to-RF Conversion (DRFC) approach, where a multi-bit D-to-A converter is merged with an image-rejection mixer into the DRFC block, as shown in Fig. 17.6. In [6], a 5.25 GHz DRFC associated with a tuned passive LC bandpass filter is proposed for WLAN and a ΔΣ modulator is used to reduce the number of bits required in the DRFC to 3 bits. The ΔΣ modulator uses a second-order pipelined MASH structure and operates with a 2.625GS/s sample rate. The circuit can deliver a maximum saturated output power of −8 dBm while consuming 187 mW. Silicon area is 0.75 mm^2.

[7] presented a 65 nm CMOS transmitter for IEEE 802.11b/g and 802.16e WLAN and WiMax standards in the 2.4–2.7 GHz band. It uses a pipelined MASH ΔΣ modulator to reduce the LSBs of the input signal, creating a 6.15 b signal (71-state word) at the RF-DAC input. The measured ACLR is 42.8/46.2 dB for 20 MHz channels at maximum power (2.6 dBm). EVM is around 2% and the transmitter offers a large 64 dB power control range. The power consumed at maximum output power is 210 mW. The core occupies a 0.35 mm^2 area.

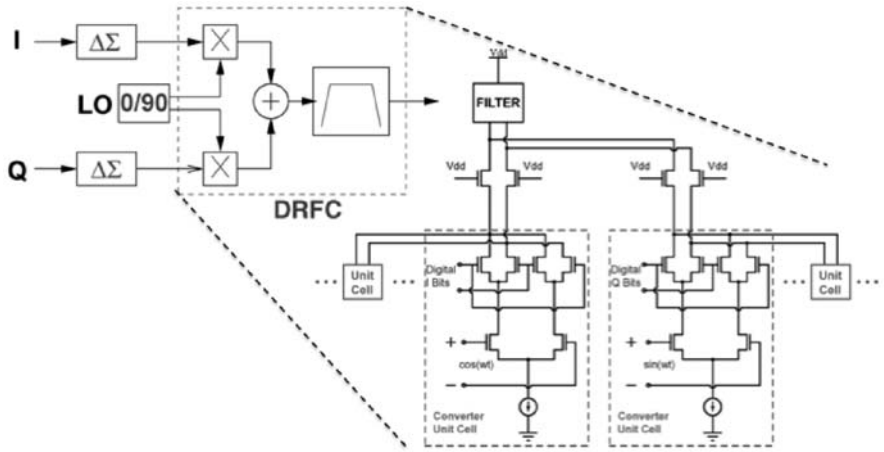


Fig. 17.6 Multi-bit ΔΣ DRFC architecture (from [6])

In these implementations, the current-mode output enables a multi-bit DAC, analog power control and good immunity to power supply noise in a differential structure. However, the current-mode output has the disadvantage of requiring a tuned on-chip matching network, thus limiting the configurability of the transmitter chain. The single bit voltage–mode approach, on the other hand, produces more quantization noise out of the considered band, but could be of high interest if efficient class-S power amplification becomes feasible. In this paper, the constraints on main signal processing blocks will be detailed to indicate the differences between these approaches in terms of complexity, performance and efficiency.

17.3 Switched Power Amplification

17.3.1 Current and Voltage-Mode Switched Power Amplifier

Switching-mode power amplifiers can theoretically achieve 100% efficiency. Current and voltage switching mode are however two really different approaches of power D/A conversion, even if the power delivered to the load is the same. In terms of power consumption, for a 1-bit switching amplifier, the voltage switching mode topology draws current from supply only for signal components inside the bandwidth as the out-of-band components are rejected by the band pass filtering stage. On the contrary, the output stage of a current switching mode topology exhibits very low out-of-band impedance thus greatly increasing power consumption and lowering power stage efficiency. Figure 17.7 shows waveforms for both switching modes. The maximum efficiency (defined as the ratio between the power delivered to the

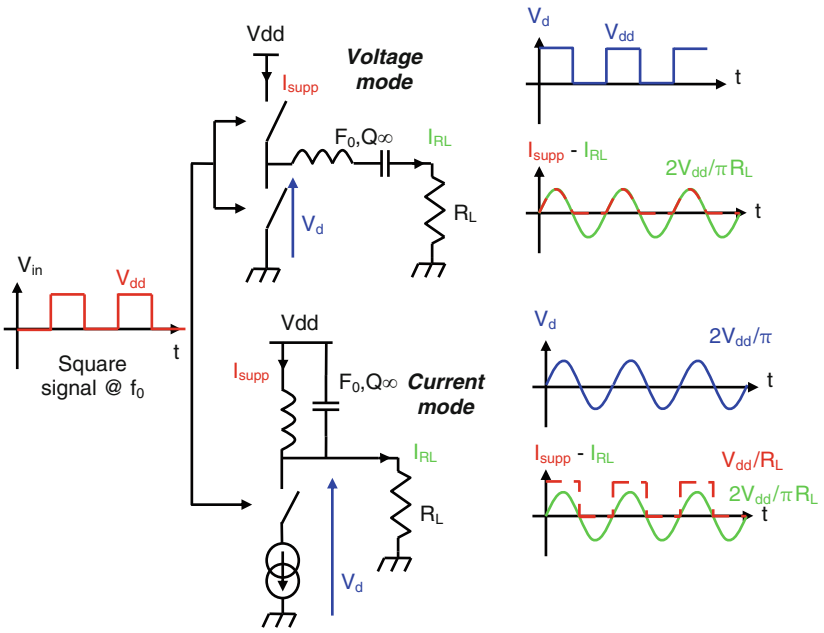


Fig. 17.7 Output waveforms for a voltage and current switching mode amplifier

Table 17.1 Efficiency comparison between voltage and current-mode switched power amplifiers

	Useful power on the load	Supply power	Theoretical Maximum efficiency
Voltage-mode	$2\ Vdd^{2}/\pi^{2}R_L$	$2\ Vdd^{2}/\pi^{2}R_L$	1
Current-mode		$Vdd^{2}/2R_L$	$4/\pi^{2}$

load and the power delivered by the power supply) for these stages is respectively 100% for the voltage mode and 40% for the current mode as stated in Table 17.1.

The easiest voltage mode switching power amplifier is a chain of scaled CMOS inverters, designed to drive a 50 Ω load under 1 V voltage supply. However, this architecture disables power control and is very sensitive to power supply noise. Note that this architecture is not suitable for multi-bit D/A conversion, because of voltage summation issues.

17.3.2 Power Combining

Downscaled CMOS technologies allow very high-speed digital processing but are not suitable for power applications, even in the range of a few dBm. The gate oxide becomes always thinner and as the breakdown voltage decreases, these technologies

require lower supply voltages. The threshold voltage does not decrease the same way, still reducing the output dynamic range. The obvious question is how to generate power with 1 V voltage supply and with very low output voltage ranges.

Nowadays, power transistors technologies (HEMT in AlGaAs or InGaP, LDMOS . . .) are expensive and are generally not compatible with CMOS processes. A fully integrated CMOS silicon solution is a very challenging task and power combining techniques are consequently a very promising research field. At low voltages, inserting an impedance matching network between the power amplifier and the load is usually necessary to increase the output power. An alternative approach is to use several power amplifiers in parallel and to combine the outputs in some way. For current output amplifiers, this is easily performed by inherent current summation on the output node. If voltage-mode amplifiers are used, then the output power may be combined using transformers [8] or transmission lines.

A transmission line-based network has been used in [9, 10] to combine the outputs of several Power DACs (PDACs). The power combiner is built with five channels driving a real load impedance R_L, representing the antenna or the input of an antenna filter. This number of channels is only restricted by the available area and can be extended to an N-channel configuration. In each channel a CMOS inverter drives a quarter wavelength transmission line with characteristic impedance Z_0. Figure 17.8 shows the topology of the five-channel power combiner.

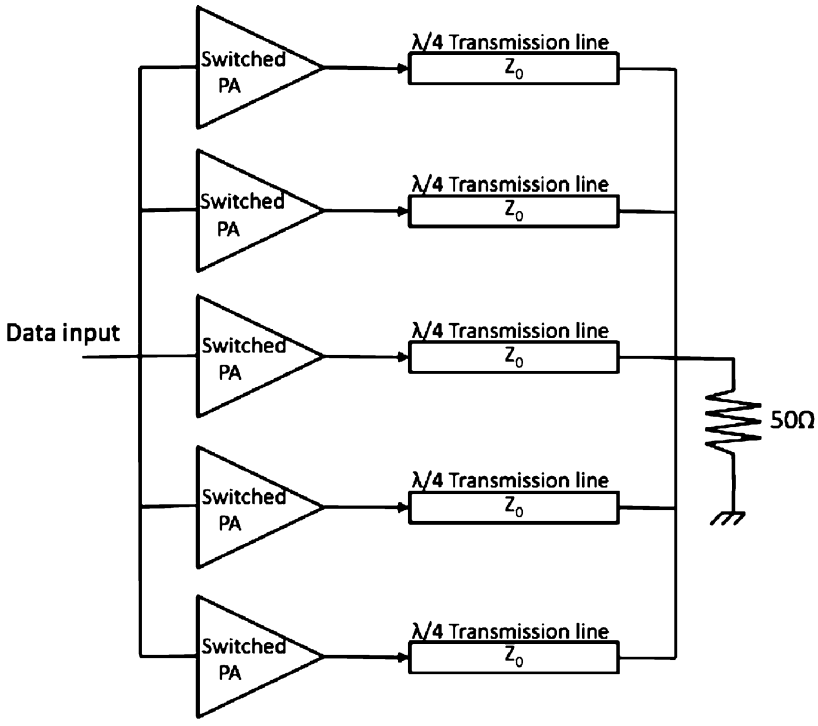


Fig. 17.8 Architecture of the five-channel power combiner

In the case of ideal transmission lines, i.e. without losses, the power gain, defined as the ratio between the output power for N channels activated and for 1 channel activated, equals to:

$$G_{p(f=f_0)} = N^2(Z_0^2/R_L + r_s)^2/(Z_0^2/R_L + Nr_s)^2 \tag{17.1}$$

where N stands for the number of activated channels and r_s for the output impedance of the PDACs. Z_0 is chosen large enough (50 Ω here) to provide a good power transfer at the expense of output power. In this case, power transfer tends towards unity while power gain is proportional to N^2.

17.4 Discussion on Digital and Mixed-Signal Blocks Implementation

17.4.1 Oversampling Up to RF Frequencies

Baseband signals have to be oversampled up to RF sampling frequencies. Poor rejection of spectral images could have a non-desirable effect on the following blocks, especially on the RF mixer. In the design of our transmitter, we have taken advantage of the large quantization noise generated by the ΔΣ modulators for greatly reducing the complexity of the oversampling stages. In fact, once moderately oversampled with good image attenuation, the input signals are delivered to the RF signal generator at a $f_s/32$ sample rate. They can then be oversampled at no additional cost with zero-order sample-and-hold interpolation, thanks to the fact that non-attenuated images situated at $f_s/32$ offsets are masked by the quantization noise. This block only uses flip-flop registers to sample the input signal at the desired sampling rate.

This compares very favorably with transmitters based on multi-bit ΔΣ modulation. Multi-bit modulators produce less out-of-band quantization noise and more baseband digital filtering attenuating the spectral images is needed. This can add a significant contribution to the total power consumption. For instance, in [7], a cascade of digital upsamplers/filters is implemented up to 1.35 GHz sampling frequencies and consumes a large portion of the total digital power.

17.4.2 Sampling Clock Synchronization on RF LO

In [6], the image cancellation is affected by non-ideal phase matching of the quadrature LO signals driving the DRFC. In fact, the sampling clock is not synchronized on the RF LO. Thus, degradation in SNR due to phase inaccuracies is reported. Synchronizing the sampling rate of the digital domain and the RF LO (used in the DAC)

is useful for exactly cancelling the image channel. In our design, the sampling clock is directly acting as the output RF LO.

Nevertheless, a major drawback in all-digital transmitters, in which the sampling clock is synchronized on the RF LO, is the fact that the ratio of the IF to the RF sampling frequencies is not necessarily integer, as one is related to the baseband chip rate and the other to the RF carrier frequency. The two ways to deal with this issue are, on one hand, to implement a non-integer sample rate conversion early in the baseband or, on the other hand, to slightly offset the RF sampling frequency to reach an integer ratio and to correct the channel center frequencies in baseband by the opposite amount. The second choice is easily implemented at no additional cost although it requires a margin on the modulator bandwidth.

17.4.3 Implementing the ΔΣ Modulator

There are a lot of ways to design ΔΣ modulators depending on the performance we want to achieve in the transmitter. There is obviously a tradeoff between performance, complexity and consumption that should be addressed according to the targeted application.

17.4.3.1 Pipelined MASH Structures

For RF transmitters, the ΔΣ modulator has to be sampled at very-high frequencies, because a high OSR is needed for achieving good SNRs for most standards. The easiest way of designing a high-speed ΔΣ modulator that operates at several GHz is a Multi-Stage Noise-Shaping (MASH) structure [6, 7], as shown in Fig. 17.9. As there

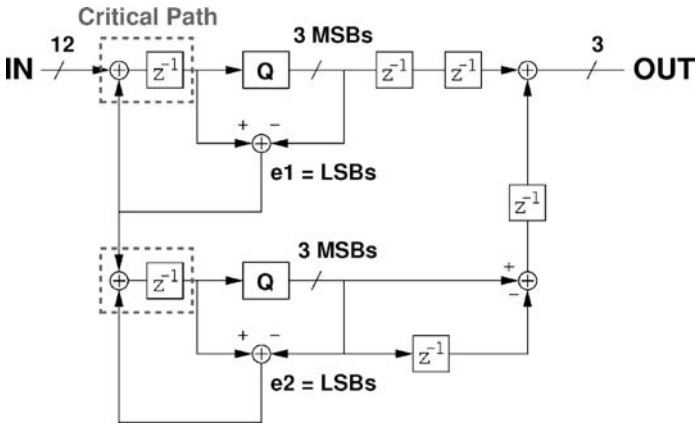


Fig. 17.9 MASH ΔΣ modulator structure used in [6]

are no feedbacks that include multiplication by negative power-of-two coefficients in this structure, it can be highly pipelined, thus reducing the critical path length, even down to 1-bit additions. This is obviously done at the expense of consumption, while increasing the number of registers.

The disadvantage of using such MASH structures lies in the fact that the transfer function for the noise cannot be engineered in a smart way. In fact, all zeros of the NTF are placed in the center of the RF band of interest (on $f = 0$ for a lowpass implementation), thus reducing the flexibility of the modulator as well as limiting its bandwidth by the rapid increase of the quantization away from the center frequency. Nevertheless, increasing the order of the modulator or increasing the OSR would still improve the performance, but once again at the expense of larger power consumption.

17.4.3.2 High-Speed ΔΣ Modulator Implementation Using Redundant Representation

Designing more flexible ΔΣ modulators requires adding complexity in the design by introducing feedbacks and other means to place poles and zeros of the transfer functions at defined frequencies and even make them programmable. A common practice is to place additional zeros on the edges of the desired bandpass, in order to obtain an almost flat noise in the band of interest. Most of the time, such structures are not suitable for pipelining. When dealing with the high-speed implementation, it appears almost impossible to fit the critical path inside the sample period. The critical path is composed of several adders for large bit-width internal signals (up to 16–20 b), where carry propagation becomes the limiting factor. A redundant representation instead of the conventional 2's complement representation has been proposed for the implementation of high-speed ΔΣ modulators [11, 12]. The chosen redundant borrow-save (BS) arithmetic enables carry-free additions by simply computing bit positions in parallel and saving carries in its representation, each bit position being doubled. Nevertheless, as the output quantizer cannot be easily computed without re-introducing a carry-propagation path, a precomputed non-exact quantization has been introduced. This output stage can be implemented with a digital (logic equations) or mixed-signal approach (current summation of bit values).

As an example, a lowpass third-order 1-bit ΔΣ modulator with optimized zero placement has been designed and measured successfully up to a 2GS/s sampling clock (Fig. 17.10) [5]. The optimized zero placement can be clearly observed at the edges of the pass-band in Fig. 17.11. A SNDR of 72 dB over a 50 MHz bandwidth has been measured at the digital mixer output. A single ΔΣ modulator only consumes 25 mW. However, the SNDR measured at the output of the chip degrades to 54 dB, due to supply voltage ringing and correlated jitter introduced in the output stage.

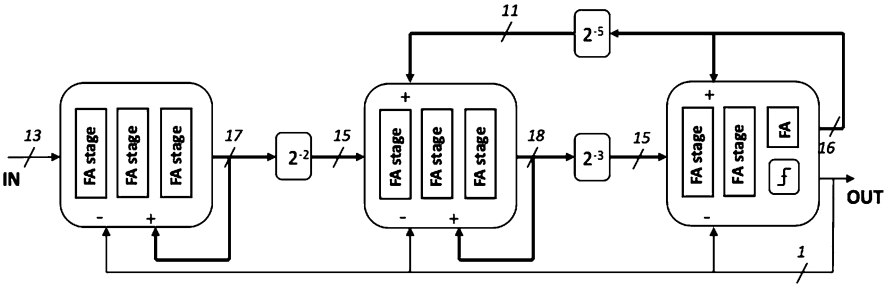


Fig. 17.10 Third-order ΔΣ modulator with optimized zeros placement (from [5])

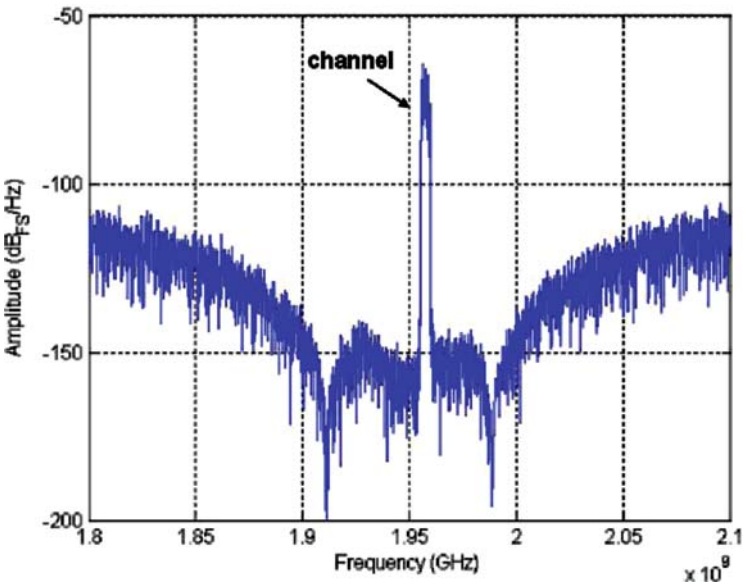


Fig. 17.11 Simulated spectrum at the output of the RF signal generator (ΔΣ modulated signal upconverted to RF frequencies)

17.4.3.3 Going Further: Adaptive Placement of Complex Poles and Zeros of the NTF

The main issue of delta-sigma based transmitters is the high amount of quantization noise generated out of band, especially if the output stream is on a single bit. This is particularly annoying because Frequency Division Duplex (FDD) standards define very low emission levels on the receive bands and other critical bands. For example, in UMTS, when transmitting in the 1.92–1.98 GHz band, the emission mask level is −183 dBm/Hz in the 2.12–2.17 GHz band, which is only 195 MHz away.

To relax the filtering requirements on the antenna filters, the previous design has opened the way to more sophisticated ΔΣ modulator designs, as it could benefit

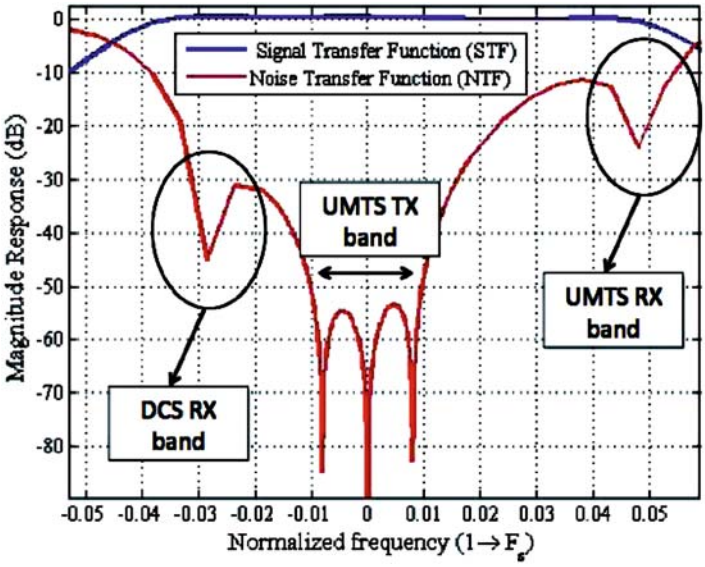


Fig. 17.12 Signal and noise transfer function for a complex lowpass ΔΣ modulator, designed for operating in the UMTS Tx band and relaxing filtering requirements on UMTS and DCS Rx bands

from continuously scaled nanometer processes. In [13], a complex IQ lowpass ΔΣ modulator proposes to increase the modulator order and place additional zeros on selected band to lower the transmitted noise. The complex approach enables another degree of freedom by moving independently the poles and zeros on both sides of the carrier frequency. A synthesis methodology has been developed to ensure stability of the whole system and keep the signal transfer function flat in the transmit band. Figure 17.12 shows an example of the noise and signal transfer function for a UMTS TX case.

Moreover, the coefficients inside the modulator could be dynamically adjusted to the targeted standard, to be able to fit different bandwidths and dynamic ranges or reduce quantization noise in unauthorized bands. By adaptively changing the transfer functions of the noise, an intelligent multi-mode transmitter can be designed for opportunistic or cognitive radios.

17.4.4 Mixer and Digital-to-Analog Conversion

Single-bit ΔΣ modulators enable the implementation of a really simple mixing stage by using a digital image-reject mixer. In the general case, two multipliers and a summer are required, all operating at the RF sampling frequency. By choosing the RF sampling frequency equal to four times the center frequency of the transmit band, the operation is however greatly simplified. In that case, the 90 degree phase shifted I and Q LO signals can be respectively represented by the $\{1, 0, -1, 0\}$ and

$\{0, 1, 0, -1\}$ digital sequences. A simple multiplexer, selecting the I channel on odd periods and the Q channel on even periods, can replace the adder in the mixer. Furthermore, the multiplications are replaced by a simple change of the sign of the digital data, eliminating multipliers entirely. The digital RF output stream is then the following sequence:

$$RF_{out} = \{I(n), Q(n+1), -I(n+2), -Q(n+3)\}, n = 0, 4, 8, 12, \ldots \quad (17.2)$$

On the other hand, in a Digital-to-RF Converter (DRFC), the mixing stage is merged with a current DAC, allowing reuse of the current, but the complexity is higher due to the multi-bit input signal. Linearity, I/Q imbalance and mismatch issues need to be addressed.

17.5 Dealing with Quantization Noise

Filtering is one of the biggest issues in digital-RF transmitters, as the digital signal processing delivers a signal that comprises a lot of quantization noise. The ideas for overcoming this problem are, first, to produce less quantization noise by, for example, using a multi-bit signal. In this case, high-efficiency voltage-mode output stages cannot be used. A second idea is to use very selective filters at the output, such as BAW filters, described in the next subsection. An intermediate approach is to relax the requirements in terms of antenna filtering by introducing a RF semi-digital FIR filtering. The most effective way should be to place the quantization noise at frequencies where it does not disturb any other communication terminals.

17.5.1 BAW Filtering

BAW (Bulk Acoustic Wave) filters have recently emerged as good candidates for high-performance RF filters supporting high power levels as required in the transmit path of a mobile transceiver. A further advantage of BAW technology is its compatibility with silicon processing. Above-IC integration of BAW resonators has already been demonstrated. As an example, the structure and the transfer-function of a differential BAW filter with 8 resonators designed for the UMTS band is shown in Fig. 17.13 [14]. It combines the advantages of the sharp transition form pass-band to stop band of the ladder structure, and the excellent stop-band rejection of the lattice structure. BAW filters are therefore ideally suited to provide the required attenuation of the out-of-band quantization noise produced by the delta-sigma modulators.

The digital RF signal generator with voltage mode output stages of [5] and a UMTS BAW filter [15] have been assembled on a PC board to demonstrate the principle. Due to series resistors that had to be inserted to adapt the BAW filter's transfer function, the insertion loss is quite high and the noise-floor in the TX band

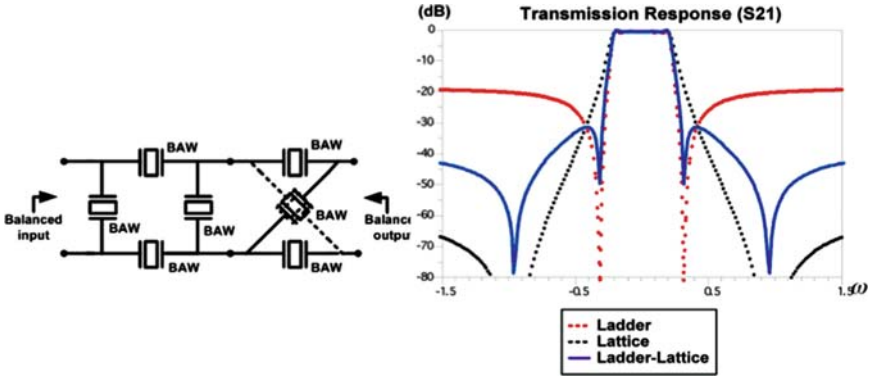


Fig. 17.13 Structure of BAW ladder-lattice UMTS band filter and corresponding transfer function [14]

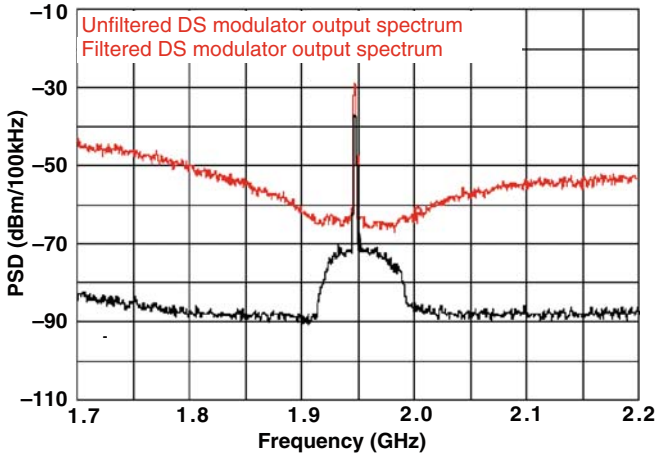


Fig. 17.14 Filtered and unfiltered ΔΣ modulator output spectra

increased above the targeted level, see Fig. 17.14. However, out-of band quantization noise has been reduced below the noise level of the spectrum analyzer in this setup.

17.5.2 RF Semi-Digital FIR Filtering

The principle of a semi-digital FIR-DAC using the output of a 1-bit ΔΣ modulator has been reported in [16]. The ΔΣ output stream feeds a digital delay line. Selected taps are summed in the analog domain by means of weighted current sources. The power combiner architecture presented in Section 17.3.2 offers the possibility to implement such a semi-digital filter at RF frequencies. The shift register is made with fast flip-flops using True Single Phase Clock (TSPC) dynamic logic. As PDACs

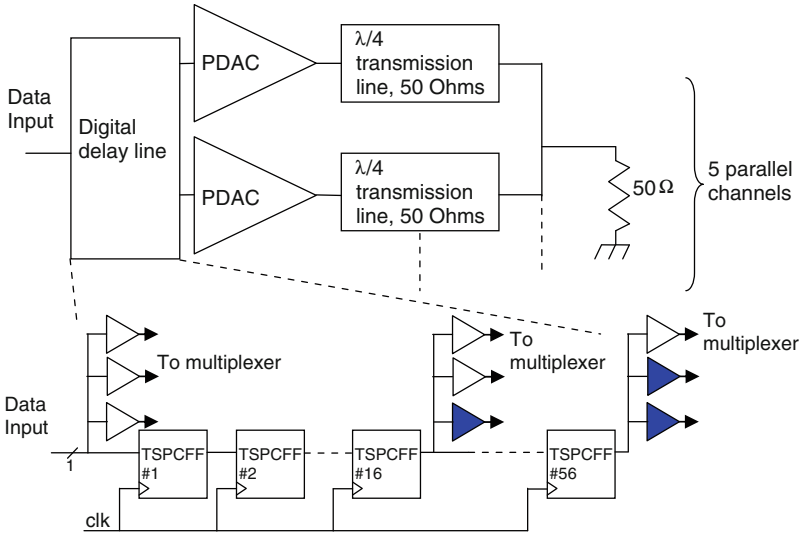


Fig. 17.15 Implementation of a digital delay line and RF FIR filter

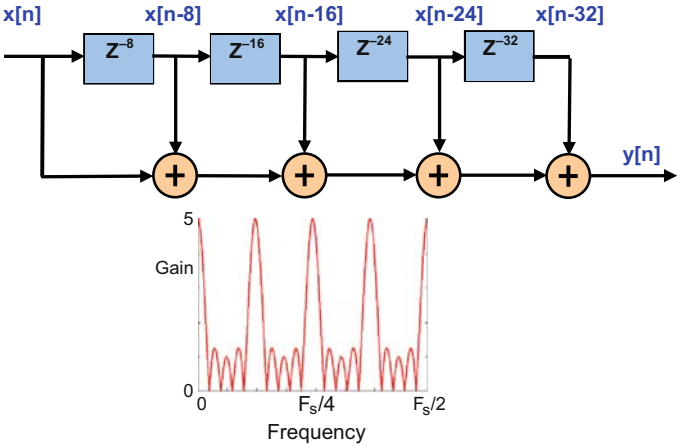


Fig. 17.16 Digital FIR filter architecture and frequency domain transfer function

operate in voltage-mode, all coefficients are equal to $+/-1$. It is indeed impossible to weight voltage sources in an inverter array. Figure 17.15 shows the global architecture including both power combiner and digital delay line.

By carefully choosing some integer coefficients and thanks to FIR properties (periodicity of transfer function), notches can be created in sensitive frequency bands, which are presently drowned by quantization noise. By simply modifying the number of delay taps, the frequency transfer function of the filter can be squeezed, resulting in a much narrower bandwidth centred on the quarter of the sampling frequency and in very precise notch locations, as presented in Fig. 17.16. As an

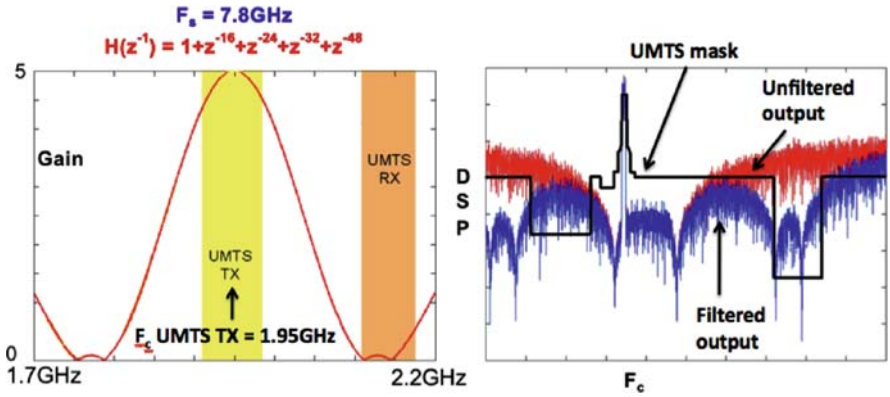


Fig. 17.17 Simulation of a semi-digital RF FIR for UMTS TX case

example, Fig. 17.17 shows a FIR frequency filter response used for the UMTS TX case in which a notch in the UMTS RX band is created (left) and a Matlab simulation on a 1-bit ΔΣ modulated signal (right, red) shows a great benefit when using this filtering technique. The filtered output spectrum is in blue and the requirements on UMTS emission mask are in black.

17.6 Conclusion

We have tried to draw a picture of multimode transmitters using ΔΣ-based signal generation and to analyze multiple ways of implementation. Directions for dealing with power amplification, signal processing and quantization noise filtering has been described but are not limited to these approaches and further research is needed to be able to build an integrated all-digital multi-mode transmitter.

Two approaches have emerged: the 1-b voltage-mode RF signal generation and the multi-bit current-mode approach. A comparison in terms of complexity, efficiency and requirements is presented in Table 17.2 to summarize the details pointed out in this paper. It is to note that power consumption and area, which are key factors, are not included as they really depend on the specific implementation. However, area and power consumption are closely dependent on complexity, efficiency and operating frequency of each block. As can be seen form Table 17.2, both approaches have advantages and drawbacks. The choice of the transmitter architecture will be driven by the targeted application, depending on factors such as output power requirements, out-of-band emission requirements, needed degree of configurability, dynamic range of the system and many other design goals.

Table 17.2 Comparison between 1-b voltage-mode and multi-bit current-mode transmitter architectures

	1-b voltage-mode	multi-bit current-mode
Output stage efficiency	+	−
Filtering requirements	−	+
On-chip tuning and calibration	+ (no tuning)	−
Output power efficiency	Depends on filter	−
Power combining	− (external power combiner needed)	+ (current summing)
Mixer complexity	+ (1-b mixer)	−
ΔΣ modulator complexity	−	+ (MASH)
ΔΣ modulator flexibility	+	− (MASH)
Oversampling blocks complexity	+	−
Power control	−	+

Acknowledgments The authors wish to acknowledge the support of the European Commission IST MOBILIS Project and A. Cathelin and D. Belot from STMicroelectronics for collaboration in the frame of the ST-IEMN joint laboratory. XLIM and CEA are also thanked for BAW design and manufacturing.

References

1. A. Jayaraman, P.F. Chen, G. Hanington, L. Larson, P. Asbeck, "Linear high-efficiency microwave power amplifiers using bandpass delta-sigma modulators," IEEE Microwave and Guided Wave Letters, vol. 8, pp. 121–123, 1998.
2. J. Keyzer, J. Hinrichs, A. Metzger, M. Iwamoto, I. Galton, P. Asbeck, "Digital generation of RF signals for wireless communications with band-pass delta-sigma modulation," IEEE MTT-S Int. Microwave Symposium Digest, vol.3, 2001, pp. 2127–2130.
3. S. R. Norsworthy, R. Schreier, G. C. Temes, "Delta-sigma data converters theory, design, and simulation," IEEE Press, 1997.
4. J. Sommarek, J. Vankka, J. Ketola, J. Lindeberg, K. Halonen, "A digital modulator with bandpass delta-sigma modulator," in Proc. IEEE ESSCIRC, 2004, pp. 159–162.
5. A. Frappé, B. Stefanelli, A. Flament, A. Kaiser, A. Cathelin, "A digital ΔΣ RF signal generator for mobile communication transmitters in 90 nm CMOS", in Proc. IEEE RFIC, 2008, pp. 13–16.
6. A. Jerng, C.G. Sodini, "A wideband ΔΣ digital-RF modulator for high data rate transmitters," IEEE J. Solid-State Circuits, vol.42, no. 8, pp.1710–1722, Aug. 2007.
7. A. Pozsgay, T. Zounes, R. Hossain, M. Boulemnakher, V. Knopik, S. Grange, "A fully digital 65 nm CMOS transmitter for the 2.4-to-2.7 GHz WiFi/WiMAX bands using 5.4 GHz ΔΣ RF DACs," IEEE ISSCC Dig. Tech. Papers, 2008, pp. 360–619.
8. I. Aoki, S.D. Kee, D.B. Rutledge, A. Hajimiri, "Distributed active transformer – a new power-combining and impedance-transformation technique", IEEE Trans. Microwave Theory and Techniques, vol. 50, no. 1, January 2002.
9. A. Flament, A. Frappé, A. Kaiser, B. Stefanelli, A. Cathelin, H. Ezzeddine, "A 1.2 GHz semi-digital reconfigurable FIR bandpass filter with passive power combiner," in Proc. IEEE ESSCIRC 2008, pp.418–421.
10. A. Flament, A. Kaiser, A. Cathelin, "Circuit intégré en particulier pour application radiofréquence multinormes et procédé correspondant de traitement d'un signal numérique radiofréquence", Patent B08–0326FR.

11. A. Frappé, A. Flament, A. Kaiser, B. Stefanelli, A. Cathelin, "Design techniques for very high speed digital delta-sigma modulators aimed at all-digital RF transmitters," in Proc. IEEE ICECS 2006, pp. 1113–1116.
12. A. Cathelin, A. Frappé, A. Kaiser, "Method for processing a digital signal in a digital delta-sigma modulator, and digital delta-sigma modulator therefore", Patent application WO 2008/102091 A3, 2008.
13. C.N. Nzeza, A. Flament, A. Frappé, A. Kaiser, A. Cathelin, J. Muller, "Reconfigurable complex digital delta-sigma modulator synthesis for digital wireless transmitters", in Proc. IEEE ECCSC, 2008, pp. 320–325.
14. E. Kerhervé, M. Aid, P. Ancey, A. Kaiser, "BAW technologies: development and applications within martina, mimosa and mobilis ist european projects", IEEE Int. Ultrasonics Symposium, Vancouver, Canada, October 2006.
15. E. Kerhervé, A. Cathelin, P. Vincent, JB. David, A. Shirakawa, "A mixed ladder-lattice BAW duplexer for W-CDMA handsets," ICECS'2007, Marrakech, 2007, December 11–14.
16. D.K. Su, B.A. Wooley, "A CMOS oversampling D/A converter with a current-mode semidigital reconstruction filter," IEEE J. Solid-State Circuits, vol. 28, no.12, December 1993, pp. 1224–1233.

Chapter 18
Switched Mode Transmitter Architectures

Henrik Sjöland, Carl Bryant, Vandana Bassoo, and Mike Faulkner

Abstract With the introduction of new cellular phone standards with increased modulation complexity and signal bandwidth, the design of efficient transmitters will be a major challenge. With this in mind a number of switched mode transmitter architectures are described, both polar and Cartesian ones. The most promising candidates use a combination of supply voltage modulation and other techniques, such as radio frequency pulse width modulation (RF PWM) or polar delta sigma modulation. This takes advantage of the high efficiency of supply voltage modulation combined with the high speed of the other techniques. Since the architectures are polar, however, significant bandwidth expansion occurs, increasing the requirements on the baseband circuits. Also for Cartesian architectures supply voltage modulation can be used to increase the efficiency, for instance in a double LINC architecture.

18.1 Introduction

The data-rates in wireless systems increase rapidly, both for short range systems like wireless LAN and for cellular systems [1]. To accommodate the higher data rates several techniques are applied. The bandwidth is increased, since everything else unchanged, the data rate is proportional to the bandwidth used. The bandwidth is limited, however, and other techniques must also be used to increase the data rate. Higher order modulation is therefore used. In the cellular standard LTE, 64-QAM is used for the highest data-rates. This allows three times higher data rate compared to QPSK. Also MIMO can be used, with several antennas used in both receiver and

H. Sjöland (✉) and C. Bryant
Lund University, Sweden

H. Sjöland
Ericsson, Lund, Sweden
e-mail: Henrik.Sjoland@eit.lth.se

V. Bassoo and M. Faulkner
Victoria University, Melbourne, Australia

A.H.M. van Roermund et al. (eds.), *Analog Circuit Design: Smart Data Converters, Filters on Chip, Multimode Transmitters*, DOI 10.1007/978-90-481-3083-2_18,
© Springer Science+Business Media B.V. 2010

transmitter, allowing multiple data streams to be transmitted simultaneously using the same frequency. The different data streams are then transmitted in different directions, and then reflected on different objects in the environment, creating different paths from transmitter to receiver.

Given this, designing an efficient transmitter for LTE cellular phones is a major challenge. The increased bandwidth is difficult to handle in pre-distortion systems and in polar transmitters, where the signal suffers from bandwidth expansion. Very high speed digital to analog converters may be needed. Also the relative duplex distances will be reduced, increasing the problems with out-of-band noise. Even worse, however, are the increased linearity requirements forced by the higher order modulation. The efficiency will be seriously degraded by having to back-off the power amplifier to accommodate the increased peak to average power ratio (PAPR), equal to 7.6 dB for 64-QAM in LTE. When using MIMO the complexity of the transmitter is increased by having several parallel transmitters. To keep the size and cost down, it is therefore highly desirable to reduce the amount of external RF filters used in each of these. This is also necessary as the number of frequency bands to support grows quickly, and a cell phone will have to support more than 10 different radio frequency bands in the near future. In addition it must also be able to handle multiple cellular standards.

To obtain highest efficiency the output stage should be operated in switched mode. The transistors are then operated as a switches; either they are closed or open. When closed, ideally they have no voltage drop, and when open they conduct no current. Since the power dissipation is voltage drop multiplied by current, there is ideally no power dissipation in the transistors. In reality, especially at high frequencies, however, there are of course losses in the transistors. Nevertheless, high efficiency can be achieved by switched mode RF power amplifiers [2].

In a transmitter for signals with complex modulation, either a polar or Cartesian architecture can be used [3]. In the base-band the signal is represented by the real part signal (I) and the imaginary part signal (Q). The Cartesian transmitter up-converts the I and Q signal to radio frequency, with quadrature local oscillator (LO) signals used in the up-conversion mixers. This means that the RF carrier of the Q signal will be 90 degrees phase shifted compared to the I carrier. The two signals are then added and fed to a power amplifier capable of handling the complex modulated RF signal. It is also possible to first amplify the two signals separately, and then add them after the power amplifiers. This description is greatly simplified, not detailing on for instance filtering. A major advantage of Cartesian architectures is that the signal is used in the same form as in the base-band. No non-linear transforms are used, avoiding bandwidth expansion. However, most switched mode architectures are polar. In a polar architecture, the signal is represented by amplitude and phase instead of real and imaginary part. This means a non-linear coordinate transformation is necessary, resulting in bandwidth expansion, see Fig. 18.1. The problems are worst when the signal passes close to the origin of the I-Q plane; the phase then makes a fast 180 degree transition with substantial high frequency content. The phase and the amplitude are also processed differently, in contrast to the symmetrical Cartesian architecture. For instance it is common

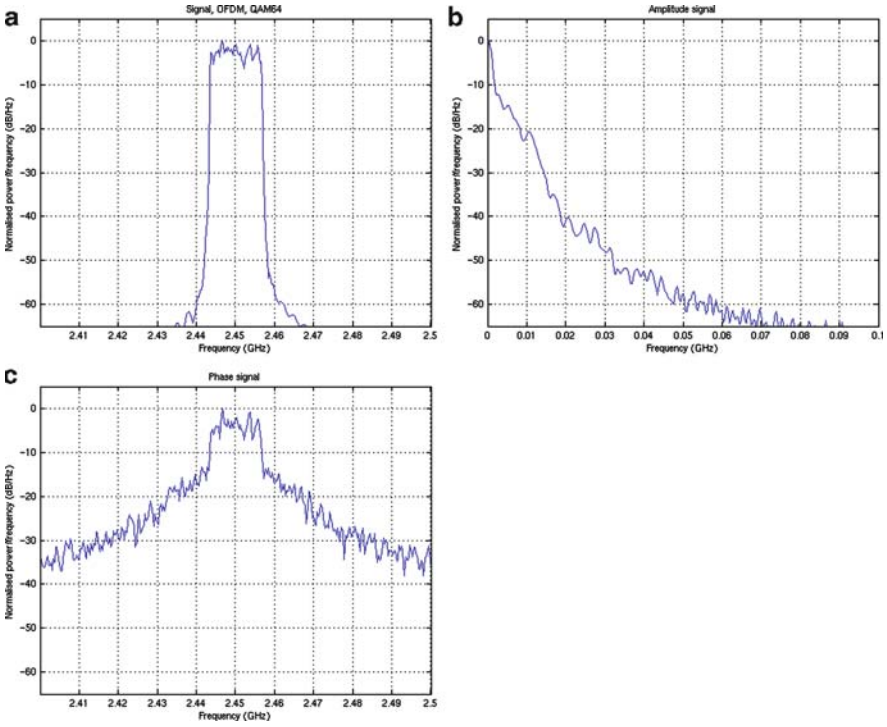


Fig. 18.1 Spectrum of 64QAM OFDM (**a**) RF signal (**b**) amplitude and (**c**) phase

that the amplitude signal is connected to a DC/DC converter controlling the output stage supply voltage. It is then very important to ensure that the phase and amplitude signal paths have equal delay. Too much delay mismatch results in the error vector magnitude (EVM) failing to meet the specification and in spurious emissions [4].

Pulse width and pulse position schemes can be implemented using time continuous or synchronous (time quantized) circuits. In the former, the pulse edges occur at any time while the latter scheme forces pulse edges on to timing grid locked to the master clock of the digital circuit. Both circuit types will be discussed.

Section 18.2 starts with an overview of switched mode RF power amplifiers and continues with envelope modulation techniques (Section 18.3). Based on this, different polar and Cartesian architectures are described with their strengths and weaknesses (Sections 18.4 and 18.5 respectively). With RF PWM identified as a key technique, then RF pulse width modulators are described, firstly using time continuous circuits (Section 18.6) and secondly with synchronous circuits (Section 18.7). Finally Section 18.8 concludes the work.

18.2 Power Amplifiers

Power amplifiers are the key transmitter building block, as they determine much of the performance and have a major influence on the architecture. They therefore serve as a good starting point.

There are different classes of operation for RF power amplifiers. First there are the linear classes; A, B, and C. As this paper focuses on switched mode operation we will not further discuss those. Instead we will discuss the switched mode classes D, E, and F. In Fig. 18.2 a simple class D output stage is illustrated.

The stage is a CMOS inverter, and the transistor output is switched either to the supply or to ground, creating a square-wave. A filter is inserted between the transistor output and the load to suppress harmonics, leaving a sinusoid at the fundamental frequency. Typically this network not only provides filtering as in the figure, but also an impedance transformation. The transistors are then often loaded by an impedance lower than that of the antenna. The amplifier can thereby deliver sufficient output power also at low supply voltages. Since the transistors operate as switches, and there are no losses in ideal inductors and capacitors, ideally the efficiency is 100%. In practice, however, there will be losses degrading the efficiency. There will be losses due to the on-resistance of the transistors, due to the losses of the LC-filter, and due to switching. Every time the stage switches, the capacitances at the input and output are charged or discharged which takes energy. The higher the switching frequency the more power is thus lost. There may also be a shoot-through current if both transistors conduct simultaneously during part of the switch event, further reducing the efficiency at high frequencies. The class-D output stage is therefore not well suited for frequencies approaching the limits of the technology. However, if the technology is fast enough compared to the operating frequency, class-D brings some important advantages.

The transistor output voltage is limited to the range between ground and supply voltage. This makes the amplifier robust to changes in load impedance, and there are less reliability issues. Furthermore, the square waveform is optimal in fundamental frequency amplitude to transistor peak voltage ratio. This means that for a certain

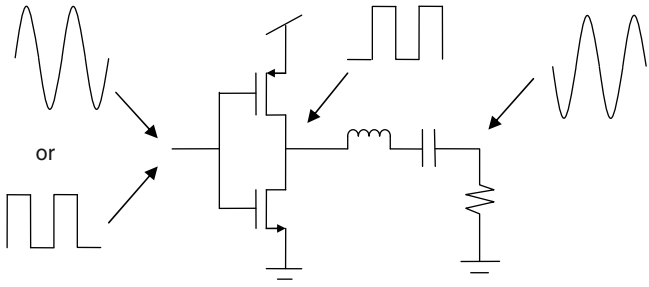


Fig. 18.2 Simple class-D output stage and signal waveforms

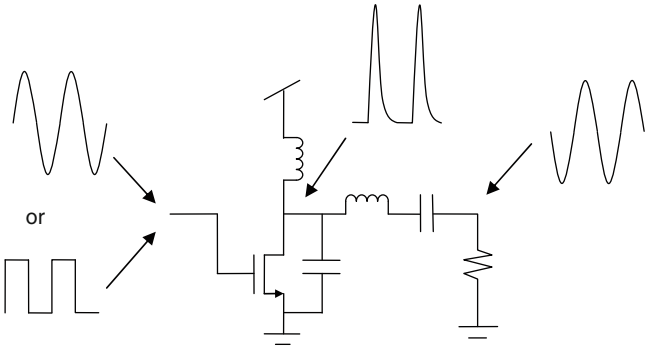


Fig. 18.3 Simple class-E output stage and signal waveforms

transistor stress, maximum output power can be generated using class-D. Last but not least, class-D supports RF pulse width modulation [5].

A simple class-E stage is depicted in Fig. 18.3.

The class-E amplifier has an improved high frequency efficiency compared to class-D. The output capacitance, which reduced the performance of class-D, is here incorporated in the output network. The transistor output voltage has a characteristic shape, formed by this network. The drain voltage is close to zero when the transistor should turn on, reducing the switching losses. The transistor turn-off, however, is still causing losses.

There are, however some disadvantages. The drain voltage is not bound between ground and supply voltage. Under normal operation the peaks of the output voltage are 3.6 times the supply voltage, and under mismatched load conditions they can be even higher. Care must therefore be taken to guarantee the reliability of the amplifier. Since the output voltage is spiky, and not square as in the class-D, the output power is not as high as the peak voltage would suggest. In other words, the fundamental amplitude to peak transistor output voltage ratio is low, and the amplifier cannot deliver as much power as class-D for a certain transistor stress. Another disadvantage is that class-E does not support RF pulse width modulation as well as class-D. The reason is that the efficient zero voltage and zero derivative voltage switching condition only occurs at one output level. It is therefore difficult to use pulse width modulation and maintain high efficiency.

The class-F amplifier solves some of the issues of the class-E, by using a more complex output network. The network shows high impedance at odd harmonics and short circuit at the even ones. This supports the waveforms of 50% duty-cycle class D. The transistor stress is therefore similar to class-D, but the efficiency at high frequencies can be improved. However, the transistor output capacitance is not included into the output network as in class-E. Furthermore, the more complicated output network is more costly and will have more losses. Since the output network shapes the waveform to 50% duty-cycle, it is not well suited for RF pulse width modulation.

18.3 Envelope Modulation Techniques

To support complex modulations it must be possible to modulate the output amplitude of a switched mode RF power amplifier. This can be done in a number of ways, for instance supply voltage modulation, RF pulse width modulation, burst modulation, and delta-sigma pulse width modulation.

Supply voltage modulation can be used with most power amplifiers. A DC–DC converter connected to the power amplifier supply voltage is then controlled by the amplitude signal. The amplitude range is dependent on the power amplifier. There is always a lowest supply voltage, below which the amplifier cannot operate. Below this supply voltage, the signal at the output will just be due to input signal leaking through. As the supply voltage is reduced towards this lower limit, depending on the amplifier, the phase shift can change considerably, and the amplitude will also depend non-linearly on the supply voltage. Some linearization may therefore be necessary. Other issues relate to the DC–DC converter. It is important that its delay is well characterized to be able to make sure that the phase and amplitude signals face equal delays. The high bandwidth of the amplitude signal is a severe problem, especially with wideband signals with large modulation depth, such as in e.g. LTE. The already high bandwidth is then expanded considerably, posing extreme requirements on DC-DC converter speed. Switched mode and linear techniques can be combined to achieve a high speed yet efficient DC–DC converter [6].

RF pulse width modulation can be used with class-D power amplifiers. The output amplitude is highest at 50% duty cycle. As the duty cycle is changed the fundamental amplitude is reduced, see Fig. 18.4.

As can be seen in the figure, also with ideal waveforms the fundamental amplitude has a non-linear dependence on the duty cycle. The fundamental amplitude, A, can be found using a Fourier series expansion:

$$A = \frac{2 \cdot Vdd \cdot \sin(\pi \cdot D)}{\pi}, \tag{18.1}$$

where D is the duty cycle and 50% corresponds to $D = 0.50$. In real circuits there will of course also be other mechanisms that degrade the linearity, and linearization is therefore necessary. Also shown in the figure is the total harmonic distortion (THD) as a function of duty cycle. As can be seen the THD is very high at low (and high) duty cycles corresponding to low output amplitudes. The ideal PWM waveform thus has a very high harmonic content at these duty cycles. This limits the amplitude range that can be achieved, since the power amplifier will not be fast enough to reproduce the harmonics accurately. At sufficiently short pulse-width the pulses will be more or less swallowed, that is the amplifier will not produce any pulse at all. The behavior of the amplifier will thus be highly non-linear and hard to predict at too short pulses. Furthermore, as the transistor drain voltage is switched all the way between supply and ground even at short pulses, the switching losses are constant and do not decrease with output amplitude. This is in contrast to supply voltage modulation, where the supply voltage is reduced at low output amplitudes,

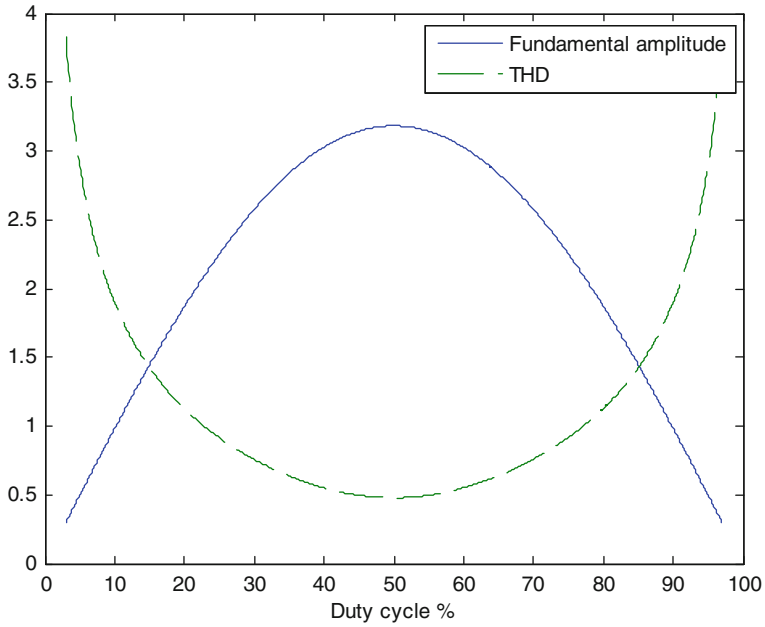


Fig. 18.4 Fundamental amplitude and total harmonic distortion vs. duty-cycle of ideal PWM waveform. (Supply voltage = 5 V)

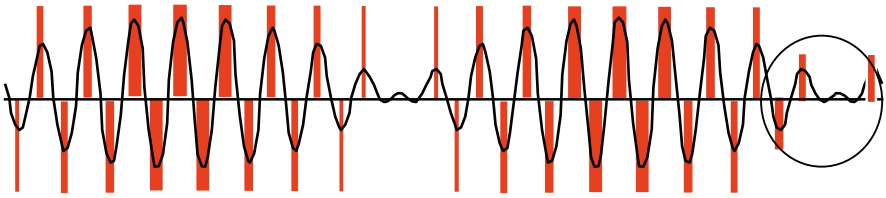


Fig. 18.5 PWM of RF carrier, with supply modulation included (circled)

resulting in fatter pulses, less switching losses and improved efficiency at low output power as shown in Fig. 18.5. An advantage of pulse-width modulation, however, is that it can handle high signal bandwidths, since in contrast to a DC-DC converter used in supply voltage modulation, there is no low-pass LC-filter involved.

In burst modulation the signal is turned on and off in signal bursts consisting of a number of RF cycles. The longer and more frequent the bursts, the higher the output power. A modulator is used to generate the activation signal. It could be a delta-sigma modulator or a pulse width modulator. Depending on the modulator different signal spectrums will appear. Using a delta sigma modulator will result in noise at frequencies surrounding the carrier. The higher the clock frequency, the further away the noise is pushed from the carrier. The loop filter shapes the noise, and using a high order filter helps reducing the noise close to the carrier. Using a pulse

width modulator will result in a less noise-like spectrum, and instead there will be strong sidebands at the PWM frequency. The noise and sidebands must be removed by an RF band-pass filters prior to the antenna, and as the trend is to remove as much RF filtering as possible, this is a strong drawback.

In delta sigma pulse width modulation the pulse width is modulated between a number of fixed values, that is it is quantized. This results in less out-of-band noise compared to the burst modulation, which uses just two pulse widths, 50% and 0% (on or off). Having discrete pulse width, the shortest pulses can be avoided, at the cost of increased noise generation requiring more band-pass filtering. The quantization in time can be used to force the pulse edges to be on a regular clock timing grid. Both pulse width and pulse position can be quantized in this way, and the advantage is that there is then compatibility with synchronous digital design.

18.4 Polar Switched Mode Architectures

The polar architectures are the most common in switched mode transmitters. Some of these are envelope elimination and restoration (EER), radio frequency pulse width modulation (RF PWM), linear amplification with non-linear components (LINC), and polar delta-sigma architectures. Also the envelope tracking (ET) architecture is polar, but it is not a switched mode architecture as it does not use a switched mode power amplifier. The hybrid EER architecture can be regarded as a combination of EER and ET. Although both hybrid EER and ET are good compromises, not being purely switched mode architectures, they are not further discussed.

In all of these architectures, except LINC, the modulated signal is represented by an amplitude signal and a phase signal at the power amplifier. The phase signal is a constant amplitude high frequency signal which is phase modulated. The phase signal is connected to the RF input of the power amplifier, which then produces a phase modulated signal at the output. To add the envelope modulation the different techniques in Section 18.3 can be used, either a single technique or a combination of several different ones.

The modulated signal can be represented by a vector in the I-Q plane, whose length and phase are modulated, that is they are functions of time. In the LINC architecture the signal is instead represented by two vectors, whose sum is equal to the modulation vector. The length of the two vectors is kept constant. At low amplitudes they are pointed at opposing directions, and at high amplitudes they are aligned, see Fig. 18.6 below.

The vectors can be generated in the digital baseband, then converted to analog and upconverted to the carrier frequency by quadrature mixers. Since the two signals have constant envelope they can be amplified by two switched mode power amplifiers. The output signals of these two power amplifiers should then be added and fed to the antenna. This is, however, more difficult than it first seems. Using a high efficiency combiner, the load impedance of the amplifiers will depend on the output power, causing errors. The architecture is very sensitive to mismatches, which

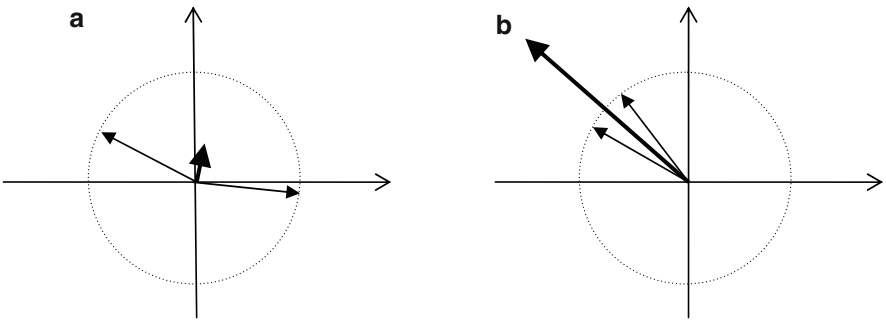


Fig. 18.6 LINC vectors (**a**) low amplitude (**b**) high amplitude

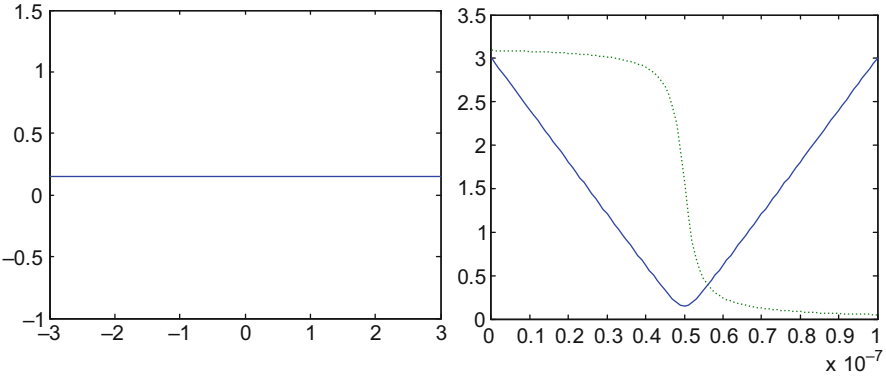


Fig. 18.7 (**a**) Signal passage in I-Q plane. (**b**) Corresponding phase and amplitude

can cause significant errors at low output amplitudes. With mismatches also the emissions to neighbouring channels may be problematic, since the bandwidth of the two signals face large expansion. Without mismatch the out of channel signals will cancel after combination, but with mismatch the cancellation will not be perfect, resulting in out of channel emissions. The bandwidth expansion problems increase with modulation depth, that is the closer the signal passes to the origin, the larger the bandwidth expansion. This is because the signals have to make a fast 180 degree phase transition containing high frequency signals at the passage of the origin. This issue is common to all polar architectures. In the polar signal representation, the phase signal has a 180 degree transition, whereas the amplitude signal has a sharp notch, see Fig. 18.7.

The sharp amplitude notch will be very difficult for a DC–DC converter in a supply voltage modulated architecture to reproduce. To reduce the bandwidth requirements of the DC–DC converter, the output voltage should not have to go all the way down in the notch. To keep the depth of the notch, a second envelope modulation technique must then be used, capable of very high signal bandwidths.

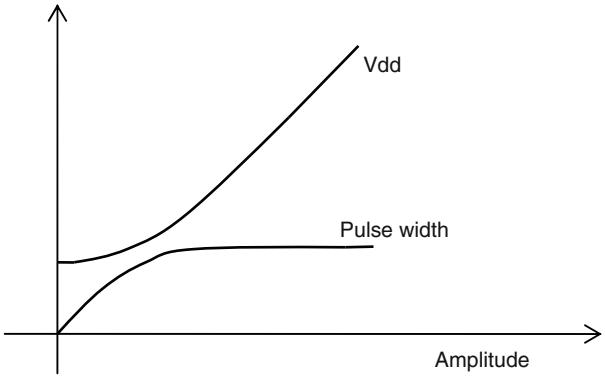


Fig. 18.8 Supply voltage and pulse width should be smooth functions of amplitude

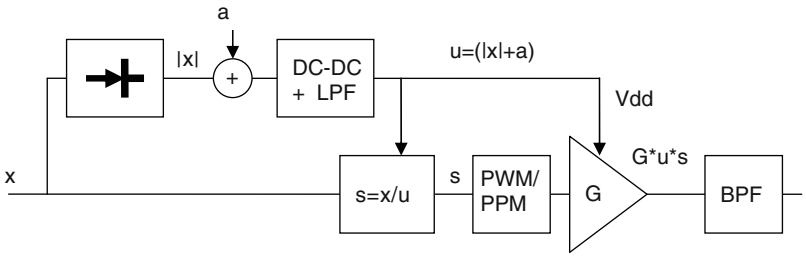


Fig. 18.9 Limited bandwidth supply modulation, with drive correction

One option is to reduce the input amplitude, making the power amplifier operate linearly instead of in switched mode, as is done in the hybrid-EER and envelope tracking (ET) architectures. However, using RF PWM the power amplifier can stay in switched mode, and a true switched mode architecture results. To avoid unnecessary bandwidth expansion, the transition between supply voltage modulation and RF PWM at low amplitudes must be gradual, see Fig. 18.8.

This combination of supply voltage and pulse width modulation is very promising. By using pulse width modulation the bandwidth requirements on the DC–DC converter can be reduced. At the same time the duty cycle is 50% at the higher supply voltages, resulting in high efficiency. When the supply voltage is lower, with lower switching losses, the duty cycle is reduced. The unfavourable situation with full supply voltage and low duty cycle, resulting in low efficiency in pure PWM architectures is thereby avoided.

The idea is for the low frequency envelope components to modulate the Vdd supply (via the DC–DC converter), and the high frequency components to modulate the pulse width modulated (PWM) input to the amplifier. The PWM's fast rise and fall times inherently enable a wide modulation bandwidth. Figure 18.9 shows a conceptual block diagram of the system.

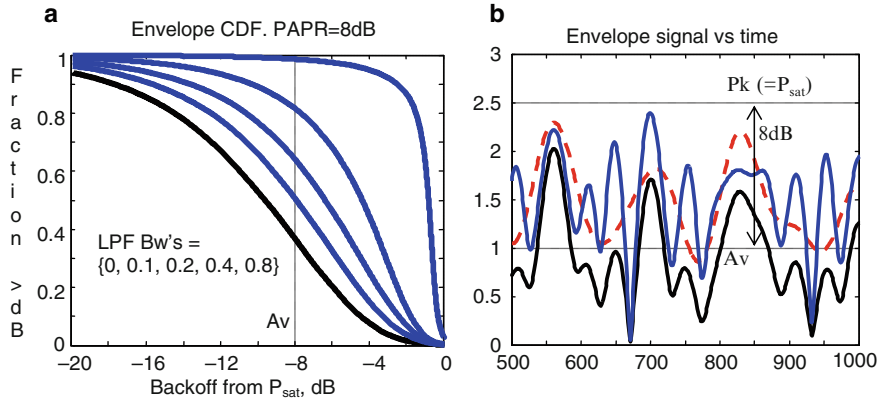


Fig. 18.10 (**a**) The probability that the drive signal amplitude, |s|, exceeds the abscissa value for LPF bandwidths of {0, 0.1, 0.2, 0.4, 0.8} channels. 0 dB is the peak saturated output power of the amplifier, P_{sat}. (**b**) Drive signal envelope, x (*black*). Corrected drive, s (*blue*) when supply signal u (*red, dotted*) is bandlimited to LPF = 0.4 channels

The envelope of the input signal (x) is found by rectification. A dc offset (a) is added to prevent clipping in the amplifier, and the resultant signal is passed to the dc-dc converter and its associated low pass smoothing filter. The compensated drive signal (s = x/u) is converted to a PWM/PPM sequence, and the resultant amplifier output becomes G * u * s = G*x, where G is the amplifier gain coefficient. A bandpass filter, BPF, removes out-of-band noise and spurs.

Having such a fast pulse width modulator also brings more benfits, since it can also be used to cancel the effect of DC–DC converter ripple. The ripple could then be measured and a compensation signal sent to the pulse width modulator. This further relaxes the DC–DC converter requirements, allowing optimization on highest efficiency rather than high bandwith and low ripple. Unfortunately the envelope signal, |x|, has a bandwidth much greater than the input signal's channel bandwidth (Fig. 18.1b). Reducing the envelope bandwidth reduces the effectiveness of the supply voltage modulation, as indicated by the cumulative distribution function (CDF) of the amplifier input in Fig. 18.10a. The input signal is an OFDM signal with PAPR = 8 dB. Ideally, with infinite LPF bandwidth, the amplifier operates at 50% duty cycle all the time and only transmits phase information (corresponds to the EER condition). As the LPF bandwidth is reduced, the drive signal (s) spends more and more time at low amplitudes, and at zero bandwidth (i.e. no Vdd modulation), the drive signal is within 8 dB of Psat for only 35% of the time (black trace). Figure 18.10b shows the time waveforms for the LPF = 0.4 channel bandwidth case. The drive signal (blue trace) now spends 82% of its time above the 8 dB backoff line, leading to reasonable pulse widths in the PWM amplifier; a good compromise between bandwidth and efficiency.

18.5 Cartesian Switched Mode Architectures

Cartesian architectures have the advantage of avoiding the bandwidth expansion associated with polar ones. In the Cartesian architectures the I and Q signals are treated separately. Each of these signals has a constant phase, and an amplitude that can be modulated with both polarities. The LINC technique can be successfully applied to such signals, since they pass exactly through the origin, and no abrupt phase transitions will then occur [7]. The LINC technique can be modified using square waves, resulting in bipolar PWM, see Fig. 18.11. In the figure one signal has been inverted compared to regular LINC, and the output is then taken differentially.

The method relies on out-phasing and therefore no short pulses needs to be generated at low amplitudes. Even zero amplitude can be safely crossed. A drawback is that the signals must be combined after the power amplifiers. In this case there are two separate LINC amplifiers, one for I and one for Q, with two power amplifiers in each. This means that the output of four power amplifiers must be combined.

Another disadvantage is that the maximum output power of the transmitter is not equal to the sum of the maximum power of the four power amplifiers. If for instance maximum power is desired at 0 degree phase, Q is equal to zero, and I is at its max. The two amplifiers in the Q branch can thereby not contribute, and just the two amplifiers in the I branch can be used. In this way just half the maximum output power can be achieved. To circumvent this, all the four amplifiers must operate in phase at large amplitudes, but in quadrature at low amplitudes. Thus a polar representation should be used at large amplitudes, and a Cartesian at low [8].

It is also possible to combine the four amplifiers into one [9]. The square-wave LINC signals are then processed by digital gates, transforming them into a single quadrature PWM signal, see Fig. 18.12.

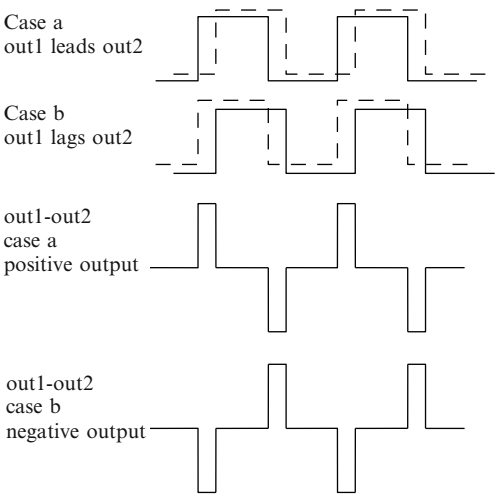


Fig. 18.11 Square-wave differential LINC signals = bipolar PWM

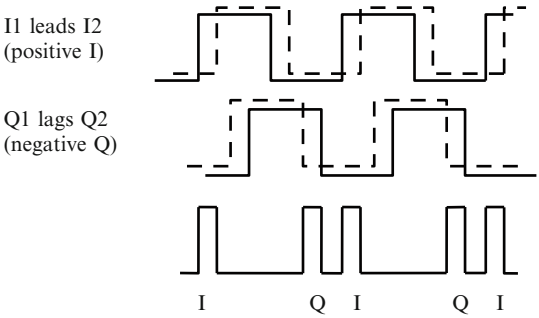


Fig. 18.12 LINC to RF PWM signals

This solution can deliver full power regardless of output phase. However, since the power amplifier must switch four times per RF cycle, instead of the minimum two, the switching losses are doubled. It is therefore not more efficient than having four amplifiers with switching losses, but only achieving the full power of two. What is gained is that the power combiner can be omitted. The drawback, however, is that very short pulses are generated close to the I and Q axes, and also short negative pulses close to maximum power.

18.6 RF Pulse Width Modulators

As have been seen RF pulse width modulation is an important technique in switched mode transmitter architectures. There are different ways to implement RF pulse width modulators. Perhaps the most simple is to just change the input bias voltage of the power amplifier, using a sinusoidal or triangular wave input signal [10, 11]. If AC coupling is used, one must be careful not to use too high an RC product. Otherwise the bandwidth will be too low for use in polar transmitters for wideband signals.

Another way is to use the square wave LINC technique, and process the two LINC signals in digital gates. Using 50% duty-cycle square wave LINC signals without any signal being inverted, the two LINC signals can be combined in a simple AND gate. The LINC signals can be generated in the digital baseband, and then converted to analog and frequency up-converted. The signals can then be converted to square-waves using limiters. The entire signal chain must be very wide-band to accommodate for the bandwidth expansion of LINC signals.

The LINC signals can also be generated in the RF domain by applying controllable delay lines, see Fig. 18.13. The phase signal, represented by a square wave, then enters two different delay lines. For maximum output amplitude the signal is subjected to equal delays in the two lines, that is the LINC signals are in phase. The amplitude signal controls the delay in the two lines in opposite directions. This means they will be more and more out-of-phase the lower amplitude is commanded, resulting in lower pulse width after the AND gate. The tunable delay lines can be

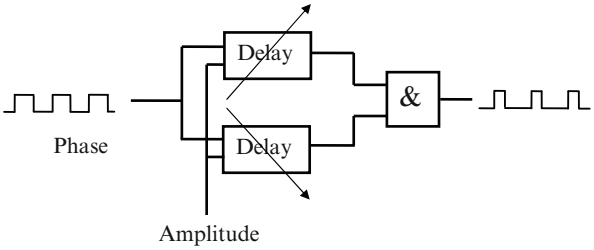


Fig. 18.13 RF PWM modulator

realized using current starved CMOS inverters. However, using long inverter chains, care must be taken not to exceed the wideband thermal noise requirements. There are also issues with linearity and robustness to PVT variations.

The linearity and robustness can be addressed using low frequency feedback [11, 12]. The DC level of the PWM output signal is then used in a feedback arrangement. As the DC level is directly proportional to the duty-cycle, when used in the feedback it can be used to set the duty cycle to the value momentarily needed. Of course there is some delay in the low-pass filter used to extract the DC value, limiting the compensation bandwidth. This is not an issue when it comes to PVT compensation, but may be for linearization. As the output amplitude is non-linear as a function of duty-cycle, see equation (18.1), the duty cycle must be set to a non-linear function of the commanded amplitude (arcsin), slightly expanding the bandwidth of the amplitude signal. The low pass filter must accommodate this.

18.7 Delta-Sigma RF Pulse Width Modulation

As previously discussed, quantization in time improves compatibility with synchronous digital circuits and can also define the minimum pulse width. Unfortunately the additional quantization noise is significant and so delta-sigma techniques are required to shape the noise away from the band of interest. They do this by subtracting the current sample's quantization error from subsequent samples [13]. The out-of-band noise, however, is enhanced and needs to be filtered by a band-pass filter at the amplifier output. The higher the delta-sigma order, the greater the band-pass filter complexity. Low order (first or second) delta-sigma systems are therefore preferred.

A key requirement of delta-sigma operation is that the sample rate, f_{sd}, must be much larger than the signal bandwidth. The input complex baseband signal must therefore be interpolated up to the delta-sigma sample rate. There are three basic architectures; Bandpass, Polar and Cartesian delta-sigma.

Bandpass Delta-Sigmas are easily obtained by replacing Z^{-1} with Z^{-2} in the low pass delta-sigma transfer function (Fig. 18.14). This moves the noise null from DC to a carrier frequency, f_c, of $f_{sd}/4$ [14]. The main problem is the coarse time

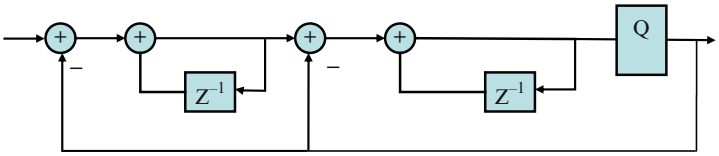


Fig. 18.14 Second order low-pass delta-sigma

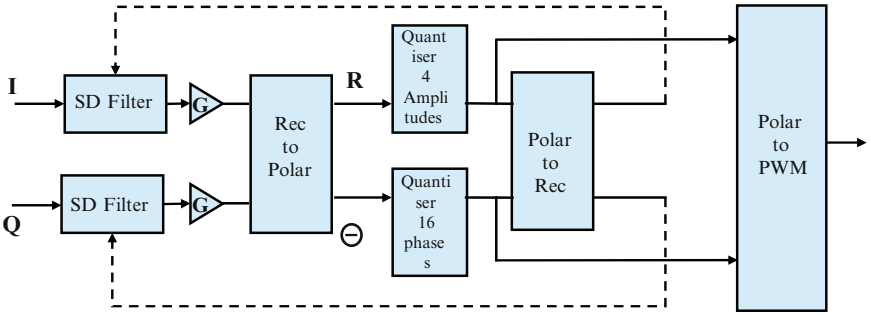


Fig. 18.15 Cartesian filtered polar D-S. Filtering uses non-bandwidth expanded IQ signal

quantization since the clock frequency f_s controlling the pulse timing grid is the same as the delta-sigma sample rate, f_{sd}. Larger oversampling ratios ($N = f_s/f_c$), for example $N = 8$, reduce quantization noise, but can lead to more than one pulse per half period of the RF carrier, which increases switching losses. The polar delta-sigma structure seeks to overcome the problem.

Polar Delta-Sigma schemes operate on a polar representation of the IQ complex baseband signal [15]. The phase and magnitude components are separately delta-sigma modulated and quantized to the available pulse widths and pulse positions. Note, there is no more than one switching pulse per period of the RF signal, potentially leading to an efficiency improvement on the 'Bandpass Delta-Sigma'.

Normally, the sigma delta filters update the pulse width and position every cycle of the carrier (or half cycle if a bridge amplifier is used). Therefore $f_{sd} = f_c$ (or $2f_c$). Unfortunately the polar components have expanded bandwidths which are likely to make it harder for the delta-sigma filters to meet the noise requirements. Also, the delta-sigma filter for the phase signal must be modified to handle the phase wrap-around. Cartesian filtering solves these problems.

Cartesian Delta-Sigma structures convert the signal to polar after the delta-sigma filters, rather than before [16]. The quantization remains in the polar domain, but must be converted back to Cartesian for the feedback signal (Fig. 18.15). In this way the filters operate in the Cartesian domain and do not see a bandwidth expanded signal. The gain terms G are normally set to 1, but can be reduced to improve efficiency (by reducing the number of switching edges) at the expense of a degraded spectrum.

The 'Polar to PWM' block in Fig. 18.15 effectively gives a further increase to the sampling rate. The quantized amplitude of the signal is transferred into a pulse width

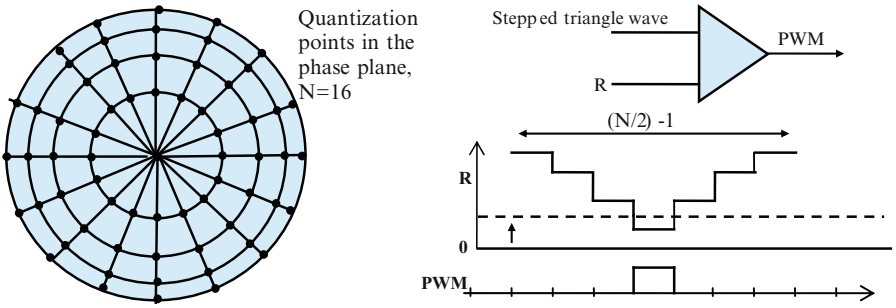


Fig. 18.16 Amplitude quantization by using a stepped triangular wave. Quantization is coarser at low amplitudes as indicated in Eq. 18.1

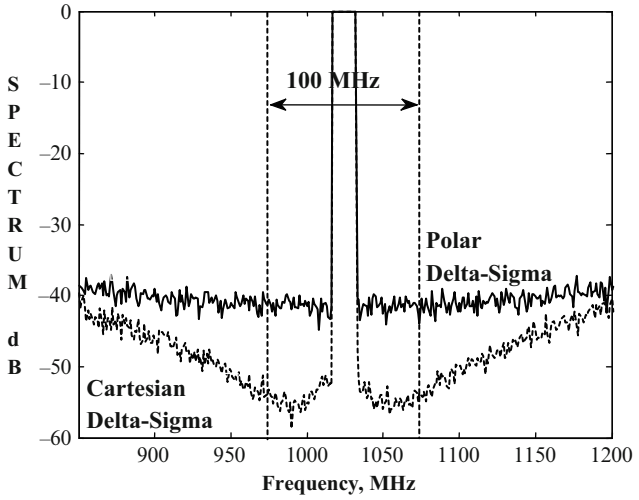


Fig. 18.17 Delta-sigma PWM spectra of a 20 MHz bandwidth OFDM signal. $F_c = 1,024$ MHz, $N = 32$

and the quantized phase of the signal into a pulse position as shown in Fig. 18.16. The number of quantization levels is related to the carrier over-sampling rate, N. Independent phase and amplitude quantization leads to a resolution of 2 π/N radians in the phase and N/4 different pulse widths. The maximum value of N is set by the technology and the carrier frequency of interest. It might be necessary to fractionally sample the clock period to get the required time resolution when carrier frequencies are high.

The Spectrum of Polar and Cartesian Delta-Sigma are compared in Fig. 18.17. Second order delta sigma filters, identical to Fig. 18.14, are used for the amplitude, phase and Cartesian signals. The Cartesian Delta-Sigma maintains adjacent channel power at less than −50 dB over a 100 MHz bandwidth.

18.8 Conclusions

There are a number of different techniques that can be applied in switched mode transceiver architectures for signals with complex modulation. The modulated amplitude can be handled for instance by varying the supply voltage in a polar architecture. This results in high efficiency. Relying on that technique alone, however, results in excessive requirements on the DC–DC converter in terms of bandwidth, and also on the power amplifier that has to operate down to zero supply voltage. Another technique is to use RF pulse width modulation, which can handle high bandwidths. Drawbacks are at low output power, where the efficiency becomes low and it is difficult to produce the narrow pulses required. These narrow pulses can be swallowed if the time axis is quantized. The resulting noise increase can best be shaped away from the band of interest using Cartesian filtered delta-sigma techniques. A combination of supply modulation and pulse width modulation techniques results in some of the most promising architectures, where the high bandwidth of RF PWM can be used to fill out the notches of the amplitude signal. The pulse width modulation can either be generated partly in the baseband using the LINC technique, or completely in the RF domain. Using the latter better facilitates DC–DC converter ripple compensation, but requires compensation for PVT variations. This can be accomplished by low frequency feedback.

References

1. E. Dahlman et al., “3G Evolution: HSPA and LTE for Mobile Broadband”, 2nd edition, Academic Press, 2008.
2. S. Cripps, “RF Power Amplifiers for Wireless Communications”, Artech House, 1999.
3. J. Groe, “Polar Transmitters for Wireless Communications”, IEEE Communications Magazine, pp. 58–63, September 2007.
4. J.F. Bercher and C. Berland, “Envelope/phase Delays Correction in an EER Radio Architecture”, In Proceedings ICECS, December 2006.
5. E. Cijvat and H. Sjöland, “Two 130 nm CMOS Class-D RF Power Amplifiers Suitable for Polar Transmitter Architectures”, In Proceedings of International Conference on Solid-State and Integrated-Circuit Technology (ICSICT), pp. 1380–1383, October 2008.
6. W-Y Chu, et al., “A 10 MHz Bandwidth, 2 mV Ripple PA Regulator for WCDMA Transmitters”, IEEE Journal of Solid-State Circuits, pp. 2809–2819, Vol. 43, No. 12, December 2008.
7. P. Dent, “Linear Amplification Systems and Methods Using More than Two Constant Length Vectors”, US 6,311,046 B1, October 30, 2001.
8. H. Sjöland, “Double LINC Switched Mode Transmitter”, US 12/139813.
9. C. Bryant, “Quadrature Pulse-Width Modulation Methods and Apparatus”, GB0822489.1
10. E. Cijvat et al., “A Comparison of Polar Transmitter Architectures Using a GaN HEMT Power Amplifier”, In Proceedings IEEE International Conference on Electronics, Circuits and Systems, pp. 1075–1078, October 2008.
11. M. Nielsen and T. Larsen, “A 2-GHz GaAs HBT RF Pulsewidth Modulator”, IEEE Transactions on Microwave Theory and Techniques, pp. 300–304, Vol. 56, No. 2, February 2008.
12. H. Sjöland, “Switched Mode Power Amplification”, WO2008002225 (A1), January 3, 2008.
13. R. Schreier and G.C. Temes, “Understanding Delta-Sigma Data Converters”, Wiley-IEEE, 2004.

14. J. Keyser, et al., “Digital Generation of RF Signals for Wireless Communications with Band-pass Delta-sigma Modulation” International Microwave Symposium Digestive, Vol. 3, pp. 2127–2130, June 2001.
15. P. Wagh and P. Midya, “High Efficiency Switched – Mode RF Power Amplifier”, 42nd Midwest Symposium on Circuits and Systems, pp. 1044–1047. Vol. 2, 1999.
16. V. Bassoo and M. Faulkner, “Sigma-Delta Digital Drive Signals for Switch-Mode Power Amplifiers”, IET Electronic Letters, Vol. 44, No. 22, October 2008.